(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068933 A2

(51) International Patent Classification[7]: G01N

(21) International Application Number: PCT/US02/06238

(22) International Filing Date: 27 February 2002 (27.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/272,248 28 February 2001 (28.02.2001) US

(71) Applicant *(for all designated States except US)*: **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(71) Applicants and
(72) Inventors: **STEVENS, Raymond, C.** [US/US]; 7076 Caminito Valverde, La Jolla, CA 92037 (US). **ORENCIA, Maria, C.** [US/US]; 778 Palmino Court, San Marcos, CA 92069 (US). **YOON, Jun, S.** [US/US]; 8868 Regents Road, # 103, San Diego, CA 92122 (US). **HANSON, Michael, A.** [US/US]; 4825 Narragansett Avenue, # 2, San Diego, CA 92107 (US).

(74) Agent: **VIKSNINS, Ann, S.**; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).

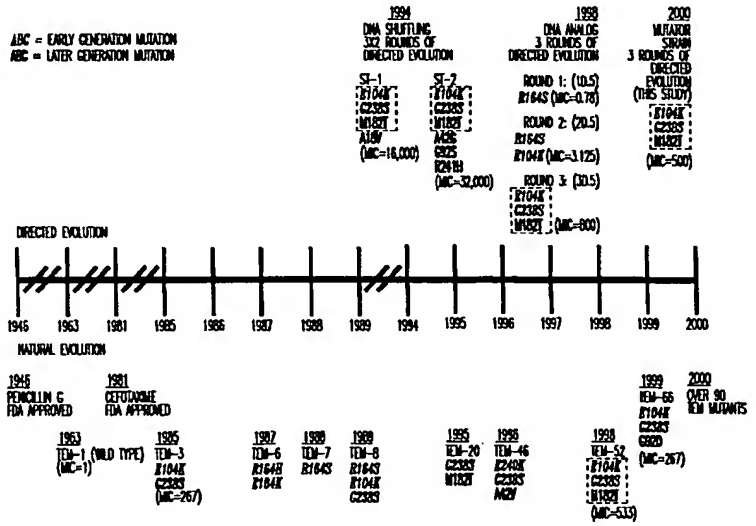
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: SMALL MOLECULE DESIGN AGAINST DRUG RESISTANT MUTANTS USING DIRECTED EVOLUTION



(57) Abstract: The present invention relates to methods for designing new drugs useful against drug-resistant bacterial cells, viruses, mammalian cells and the like. The method involves identifying a target protein of the drug, selecting for drug-resistant variants that have an altered target protein (variant protein) by directed evolution, determining the three dimensional structure of the target and variant proteins and designing a new drug that can be effective against at least one drug-resistant variant. The present invention can be used to predict future mutations that lead to drug resistance and the type of drugs that are effective to combat such resistance.

WO 02/068933 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SMALL MOLECULE DESIGN AGAINST DRUG RESISTANT MUTANTS USING DIRECTED EVOLUTION

FIELD OF INVENTION

The present invention combines the fields of directed evolution and rational drug design through the use of three-dimensional structural analysis to provide methods for identifying molecules that can effectively combat drug resistance.

BACKGROUND OF THE INVENTION

Antibiotic resistance is gaining in importance as a medical problem because more microorganisms are becoming resistant to greater numbers of antibiotics. For example, some strains of *Staphylococcus aureus* have developed resistance to several distinct antibiotics and are now described as having a "multi-resistant" phenotype. Unfortunately, due to the extensive use of antibiotics, and sometimes potential misuse of antibiotics, such as the self-prescription of ciprofloxacin (CIPRO) prevalent in the Fall of 2001, such multi-resistant strains of microorganisms are now found in the general population and within hospitals. Hospital-acquired bacterial infections occur in 5% of patients admitted to the hospital (about 2 million patients per year in the United States).

"New" antibiotics are often structurally related to or derived from a previous generation of antibiotics. For example, cephalosporin is structurally related to penicillin. While these structural analogs of known antibiotics can be successful for a time, reports of resistance become prevalent as the new antibiotic becomes widely used.

The progression of β-lactam antibiotic resistance illustrates this problem. Since the introduction of penicillin in the 1940's, microbes have evolved resistance to β-lactam antibiotics, typically by acquired or enhanced catabolism of these drugs through the enzyme β-lactamase found in many bacterial strains. While researchers have made successive changes to the basic β-lactam antibiotic structure in order to make new antibiotics that are more resistant to degradation by β-lactamases, microorganisms have responded by producing different types or larger amounts of β-lactamases. Such resistance is particularly problematic for hospital-acquired infections. A common mechanism for maintenance and transmission of β-lactam antibiotic-resistance is via plasmids that can readily amplify. When β-lactam antibiotics are routinely prescribed, as often occurs in hospitals, a small population of bacteria that are highly resistant to such antibiotics can propagate and spread throughout the hospital. This problem has

been addressed by the production of new generations of antibiotics that are usually more toxic than their predecessors, and which often cannot be administered to patients in a convenient manner. For example, vancomycin, which is often called the "antibiotic of last resort," has so much toxicity that its use is restricted to the most dangerous, multiply drug-resistant, infections.

Another therapeutic approach to avoiding antibiotic resistance has been to use β-lactamase inhibitors. Unfortunately, although the different classes of β-lactamases have overlapping substrate specificities, they have evolved differently and have distinct amino acid sequences and different three-dimensional structures. Thus, currently available β-lactamase inhibitors often have widely varying efficacies for the different β-lactamase enzymes. Moreover, even when the primary amino acid sequences of various β-lactamases are known, such knowledge does not allow researchers to predict the structure of successful β-lactamase inhibitors. For example, individual mutations at position 164 and 238 of TEM class β-lactamases occur in 35% and 32% of all antibiotic-resistant mutants, respectively, but rarely in combination. An understanding of the three-dimensional structure of TEM β-lactamase, including mutants of TEM containing mutations at position 164 or 238, is needed to understand how these mutations lead to antibiotic resistance and how to best combat such resistance. Similarly, the primary amino acid sequence information for known β-lactamase mutants does not necessarily provide information about other types of mutations that are likely to occur in the future. Accordingly, new approaches are needed for the design of antibiotics and inhibitors of antibiotic resistance that incorporate an understanding of the three-dimensional structures of the target microbial proteins involved.

Virus and cancer resistance is also a recurring problem for the medical community. The occurrence of drug resistance to new and old drugs is increasing. As therapeutic agents are used, the targets of their action evolve to develop resistance against them, lowering the probability for successful treatment of a large number of disorders involving microbial or viral infections and neoplastic transformations.

For example, RNA viral diseases are responsible for the vast majority of morbidity and mortality for viral diseases in humans. Such diseases include AIDS, hepatitis, rhinovirus infections of the respiratory tract, flu, measles, polio and others. Acquired Immune Deficiency Syndrome ("AIDS") is a fatal human disease that has recently grown to epidemic proportions. Epidemiological evidence indicates that this disease is caused by the human immunodeficiency viruses, HIV-1 or HIV-2. HIV is particularly difficult to eradicate for several

reasons. For example, HIV permanently incorporates its genetic material into the genome of infected cells, it replicates (Perelson, A. S., Science 271:1582 (1966)) and it mutates (Larder, B. A., Science 246:1155 (1989)) at an exceptionally high rates and thereby avoids immune inactivation, and it specifically infects and destroys the very immune system components most critical for controlling the infection. There is currently no effective vaccine for prevention of infection, and aside from the recent success for combination therapies involving protease inhibitors, few treatments for this fatal disease exist. Furthermore, the virus rapidly develops mutations conferring resistance against all chemotherapeutic agents tested to date. See U.S. Patent No. 6,063,628, herein incorporated by reference.

All currently approved anti-AIDS drugs are designed to inhibit either the human immunodeficiency virus reverse transcriptase (HIV RT) or the virally-encoded protease or a combination thereof. Chemicals directed against HIV RT are either deoxynucleoside analogs that terminate HIV DNA synthesis or non-nucleoside analogs that inhibit the reverse transcriptase (Larder, B. A., J. Gen. Virol. 75:951 (1994)). The guiding concept of all current AIDS therapies is to prevent further replication of the virus by directing drugs to interfere with the production or maturation of viral-encoded proteins. Unfortunately, the therapeutic effectiveness of these drugs is generally rendered obsolete by the emergence of resistant viral mutations. It has been estimated that the mutation rate of HIV is one million times greater than the mutation rate of human cells. Recently, combinations of drugs have been shown to be more effective in reducing the viral load in individuals and it is hoped that this reduction in viral load will result in prolongation of life (Ho, D., Science 272:1125 (1996)). However, the reduction in circulating virus may merely delay, rather than prevent, the development of drug resistant mutants in an infected individual.

Drug-resistant mutations also lead to difficulties in treating cancer. For instance, it is believed that many solid tumors, such as breast cancer, progress from initiation to metastasis through the accumulation of several genetic aberrations. *See*, Smith *et al.*, Breast Cancer Res. Treat., 18 Suppl. 1: S 514 (1991); van de Vijver and Nusse, Biochim. Biophys. Acta, 1072: 33-50 (1991); Sato *et al.*, Cancer Res., 50: 7184-7189 (1990). Such genetic aberrations, as they accumulate, may confer proliferative advantages, genetic instability and the attendant ability to evolve drug resistance rapidly, as well as enhance angiogenesis, proteolysis and metastasis. The genetic aberrations may affect either recessive "tumor suppressor genes" or dominantly-acting oncogenes. Deletions and recombination leading to loss of heterozygosity are believed to

play a major role in tumor progression by uncovering mutated tumor suppressor alleles. See U.S. Patent 5,965,362, herein incorporated by reference.

Mutation also plays a role in drug resistant leukemic forms of cancer. According to Wheldon and colleagues (Br. J. Cancer Suppl., 18: S13-9 (1992)), one mechanism for early onset leukemogenesis is the occurrence of a high cellular mutation rate. This mechanism is thought to enable both the rapid evolution of leukemic cells as well as the early emergence of drug-resistant variants. Upon application of chemotherapy, variant cancerous cells that are resistant to the chemotherapeutic agent administered are more likely to survive, and may eventually dominate, leading to an aggressive, non-responsive malignancy (see also Ling *et al.*, Cancer Metastasis Rev. 4(2): 173-92 (1985)). In another example, Zhao and colleagues identified four murine leukemia cell lines resistant to 5,10-dideazatetrahydrofolate (DDATHF) and other folate analogs by chemical mutagenesis followed by DDATHF selective pressure (J. Biol. Chem. 275 (34): 26599-26606 (August, 2000)). Each of the clones contained mutations in each allele of the folypolyglutamate synthetase (FPGS) gene, typically in portions of the gene coding for predicted ligand binding sites that have very high homology across species from bacteria to man. In each of the clones, the two alleles of the FPGS gene mutated independently, suggesting that more than one drug resistant variant of the same native protein having different molecular structures may be expressed in a patient simultaneously.

Accordingly, new methods are needed for designing effective drugs to combat drug resistance.

SUMMARY OF THE INVENTION

The present invention provides methods for identifying new drugs and potential inhibitors and modulators of drug-resistant variants of a target protein of a drug of interest. A drug-resistant variant according to the invention has at least one mutation resulting in a structural change, an activity change or a stability change as compared to the target protein. Such variants would include natural variants such as those encountered in the clinic, but preferably variants are selected by directed evolution methodology.

The invention therefore provides a method for identifying a new drug effective against at least one drug-resistant variant of a target protein, comprising: selecting at least one drug-resistant variant of a target protein in the presence of a drug; obtaining the three dimensional structure of at least one variant site of interaction with the drug; obtaining the three dimensional structure of the target protein site of interaction with the drug; comparing the three

dimensional structures of the target and variant sites of interaction to identify structural similarities and differences; and designing a new drug to interact with at least one of the variants by using the structural similarities or differences. The three dimensional structures of the variants and target proteins disclosed herein can be obtained by available procedures, for example, by using x-ray crystallography, NMR, or molecular modeling.

The method may further comprise contacting a new drug with one or more drug-resistant variants or with the target protein to test the ability of the new drug to interact with the drug-resistant variants and/or the target protein. An interaction may be indicated by assaying the function of the drug-resistant variant(s) or the target protein in the presence of the new drug, or by co-crystallizing the new drug and the drug-resistant variant(s) or target protein or fragment thereof. New drugs may be selected for their ability to interact with both the drug-resistant variant(s) and the target protein. Alternatively, new drugs may be selected for their ability to interact with regions of drug-resistant variant(s) that are structurally different than said target protein. New drugs would also include those selected for their ability to inhibit the folding of one or more regions of said drug-resistant variant(s) or said target protein. New drugs may be selected from a database of small molecules or ligands, and may include a combination of molecules or entities. Preferably, new drugs are designed using any one or more of the methods selected from the group consisting of structure based drug design (SBDD), virtual ligand screening, ligand screening, high-throughput screening and ultra high-throughput screening. The drug or the new drug can an antibiotic, an enzyme inhibitor, a protein inhibitor, an antibiotic resistance inhibitor, a herbicide resistance inhibitor, an insect repellant, an insecticide resistance inhibitor, a viral drug resistance inhibitor, a chemotherapeutic resistance inhibitor, a modulator, an antagonist, an agonist, an effector, a ligand, an antibody, an antibody fragment, a peptide or a small molecule.

Proteins to be targeted by the disclosed methods include, for example, bacterial proteins, viral proteins, microbial proteins, fungal proteins, insect proteins, cellular proteins contributing to the development of cancer and cellular proteins involved in inflammation or autoimmune disorders or diseases. Such proteins would include enzymes, receptors, transporters, ion channels, transcription factors, DNA polymerases, RNA polymerases, ribosomal proteins, kinases, receptors, tumor suppressor proteins, cell cycle proteins and DNA

binding proteins, to name only a few. A preferred target protein is Beta-lactamase.

Also encompassed in the present invention are methods for preparing libraries representing drug resistance evolution profiles of target proteins or target protein fragments. Such a method may comprise the steps of (a) generating a first population of sequence or structure variants of said target protein by one or more rounds of a first method of directed evolution in the presence of said drug; (b) generating at least one other population of sequence or structure variants by one or more further rounds of said first method of directed evolution or by at least one other method of directed evolution or by obtaining a library of natural variant isolates; (c) comparing sequences or structures or fragments of said variants in said first population to the sequences or structures or fragments of said variants in said at least one other population to identify variants having unique sequences or structures; and (d) preparing a library comprising said unique sequence variants that represents a drug resistance evolution profile of said target protein or target protein fragment. Such libraries may be composed of drug resistant variants that are profiled either by their primary amino acid sequences or by their three-dimensional structures. Libraries may be screened to identify potential inhibitors or modulators as disclosed herein, or may be analyzed to identify regions in a target protein that may be structurally altered during the course of a given drug treatment so as to convey resistance to a given drug.

The invention also provides a data structure for storing information on directed evolution of a series of drug resistant variants comprising: a series of separate variant data fields, each variant data field comprising each x, y and z atomic coordinate of each drug resistant variant; a target protein data field comprising each x, y and z atomic coordinate of the target protein; a structural difference data field comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein; and a series of drug resistance data fields, each drug resistance data field comprising each x, y and z atomic coordinate of each drug that is resistant to a drug resistance variant.

The invention further provides a computer-assisted method for identifying a new drug effective against at least one drug-resistant variant of a target protein, comprising: obtaining the three dimensional structure of at least one variant site of interaction with the drug by using computer processing to generate x, y and z coordinates for each atom in the variant site of interaction; obtaining the three dimensional structure of the target protein site of interaction with the drug by using computer processing to generate an x, y and z coordinates for each atom in the target protein site of interaction; measuring the differences in distance between the x, y and z coordinates for each atom in the variant and target protein sites of interaction; designing a new drug to interact with the site of interaction of at least one of the variants by using computer processing to fit a library of potential new drugs into the variant and target protein sites of interaction and determining which potential new drug fits into the variant site of interaction better than into the target protein site of interaction.

The invention also provides an article of manufacture having instructions stored on it that cause a computing system to identify a new drug effective against at least one drug-resistant variant of a target protein, the instructions comprising: instructions for comparing a series of separate variant data fields with a target protein data field to generate a structural difference data field comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein, wherein each variant data field within the series of separate variant data fields comprises each x, y and z atomic coordinate of each drug resistant variant, and wherein a target protein data field comprises each x, y and z atomic coordinate of the target protein; instructions for generating a new drug data field from the structural difference data field and a small molecule data field, wherein the small molecule data field comprises x, y and z atomic coordinates for a series of small molecules and wherein the new drug data field comprises small molecules that can optimally fit within the site of interaction of the drug-resistant variant. New drugs identified using the article of manufacture are also provided by the invention.

The invention further provides a method for identifying a new antibiotic effective against at least one antibiotic-resistant of bacteria, comprising:

selecting at least one antibiotic-resistant strain of bacteria by exposing the bacteria to a selecting antibiotic; obtaining the three dimensional structure of a target protein site of interaction with the antibiotic, wherein the target protein is within the bacteria and wherein the antibiotic has its effect through the target protein; obtaining the three dimensional structure of a target protein site of interaction with the antibiotic, wherein the target protein is within the bacteria and wherein the antibiotic has its effect through the target protein; obtaining the three dimensional structure of a variant protein site of interaction with the antibiotic, wherein the variant protein is within at least one antibiotic-resistant strain of bacteria selected in step (a) and wherein the variant protein is resistant to the selecting antibiotic; comparing the three dimensional structures of the target and variant sites of interaction to identify structural similarities and differences; and designing a new drug to interact with at least one of the variants by using the structural similarities or differences.

The invention also provides a system that includes a computer system; memory accessible to the computer system, a series of data structures stored in the memory and one or more software components operable on the computer system. A computer system is any type of device having at least a processor and accessible memory and associated networks and peripherals, such as local area networks (LANS) and printers. See e.g., Fig. 7. Examples of computers include personal computers, laptops, supercomputers, and any other kind of computer. The computer system is capable of operating according to instructions, software components, and data structures.

In particular, the invention provides a system comprising a computer system; memory accessible to the computer system; a first data structure stored in the memory comprising three dimensional structural data of a target protein; a second data structure stored in the memory comprising a first population of three dimensional structural data for a series of variants of the target protein, wherein the variants were generated by one or more rounds of directed evolution of the target protein in the presence of a drug; a third data structure stored in the memory comprising a second population of three dimensional structural data for a series of second generation variants of a first variant, wherein the second generation of variants is generated by one or more rounds of directed evolution

of the first variant in the presence of a drug; a fourth data structure comprising three dimensional structural data for a population of potential new drugs comprising antibiotics, enzyme inhibitors, protein inhibitors, antibiotic resistance inhibitors, herbicide resistance inhibitors, insect repellants, insecticide resistance inhibitors, viral drug resistance inhibitors, chemotherapeutic resistance inhibitors, modulators, antagonists, agonists, effectors, ligands, antibodies, antibody fragments, peptides or small molecules; a first software component operable on the computer system to compare the first data structure with the second or third data structure and to identify three dimensional structural variants that represent a drug resistance evolution profile of said target protein; a second software component operable on the computer system to compare the fourth data structure with the drug resistance evolution profile of said target protein and thereby to design a new drug effective against at least one drug resistance variant.

DESCRIPTION OF THE DRAWINGS

Fig. 1 is a graph depicting the frequency of mutations obtained depending on the methodology employed, and illustrates that directed evolution can identify the same β-lactamase extended-spectrum antibiotic resistant mutations as are observed in nature (clinical settings). Four mutation methods are compared: natural mutations, use of DNA analogs, use of gene shuffling, and use of a mutator strain. The arrows indicate which mutations are ESBL mutations.

Fig. 2 is a time-line illustrating trends that have been observed in β-lactamase mutation combinations and the order of mutation appearance. Early generation mutations are in italics whereas later generation mutations are in bold. *See* Orencia et al. Nature Structural. Biology 8(3):238-42 (2001); Stemmer, 370 Nature 389-91 (1984); Zaccoic et al., 285 J. Mol. Biol. 775-83 (1999); Jacoby et al., http://www.tabey.org/studies/webt.htm (2000); Poyart et al., 42 Antimicrob. Agents Chemother. 108-113 (1998); Medeiros, Clin. Invest. Dis. 24(Suppl. 1): 319-45 (1997).

Fig. 3 provides alternative views of the TEM-52 three-dimensional structure. Fig. 3A pinpoints the E104K/M182T/G238S combination of mutations (identified by numbers 104, 182 and 238, respectively) and the loop

mutations at positions 42, 92 and 241. Catalytic residue positions are at positions 70, 166 and 182 and the B3 β-strand is the strand running upward from residue 238. Figs. 3B and 3C provide an overlay stereoview of wild type TEM-1 (darker) and TEM-52 (lighter) illustrating movement of loops 238–243, 267–271 and 40–43. Figs. 3D and 3E provide a stereoview of the electron density for the TEM-52 active site. See Orencia et al. Nature Structural. Biology 8(3):238-42 (2001).

Fig. 4 depicts the structure of the global suppressor mutation M182T found in TEM-52. Fig. 4a and 4b provide alternative views of TEM-52, illustrating the interactions relevant to the Thr 182 mutation. See Orencia et al. Nature Structural. Biology 8(3):238-42 (2001).

Figure 5 is a block diagram of one embodiment of a data structure **500**.

Figure 6 is a block diagram of one embodiment of a computer-assisted method **600** for identifying a new drug effective against at least one drug-resistant variant of a target protein.

Figure 7 is a block diagram of an example system **700** involving a computer system **702** for practicing various embodiments of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention utilizes a combination of directed or natural evolution and ligand screening and/or structural analysis to provide insight into the changes in protein conformation that lead to drug-resistance, as well as provide insight into the design of compounds, *i.e.*, antibiotics and inhibitors, capable of combating drug-resistance by interacting with the enzymes and proteins responsible for drug-resistance.

According to the invention, the design and selection of drug-resistant mutant proteins or enzymes or substrates can provide clues as to the molecular basis for drug resistance. The fields of directed evolution and rational design based on three-dimensional structure have been considered to be separate fields, with the proponents of each field being quick to dismiss the utility of the other. For instance, in a recent article by Tobin and others, site-directed mutagenesis of target residues based on structural data is described as resource- and labor-intensive, as well as time-consuming, with the authors stressing that evolution of new protein function may be studied in the absence of structural knowledge (Current Opinion in Structural Biol. 10: 421-27 (2000)). In contrast, in WO0159066, "Protein Design Automation for Protein Libraries," Dahiyat stresses the utility of computer modeling based on three-dimensional structure to

identify small molecule ligands, but dismisses directed evolution because it ignores all structural and biophysical knowledge of proteins.

The present invention goes further than any of the techniques of the prior art, in that it uses directed evolution or natural selection to obtain drug-resistant mutants, and then analyzes the three-dimensional structure of such mutant proteins or enzymes, *etc.*, to predict the variants that will arise during the course of a drug treatment or that exist prior to treatment and are able to survive and dominate once a drug is administered. While others may have used either a directed evolution or structural approach, they have not realized that the combination of directed evolution and structural analysis leads to a better understanding of protein function. Moreover, none have taken such an approach and used the data accumulated to design new drugs and protein inhibitors that may be used to treat patients whose disease has become non-responsive due to drug-resistance.

The invention allows the practitioner to predict and profile the evolution of resistance within target proteins in response to administered drugs. In addition, the invention provides methods for designing antibiotics, antiviral compounds, inhibitors, and small molecules against native and evolved targets. Such target proteins include bacterial and viral proteins and proteins involved in the development of cancer or immune disorders. In general, the present methods involve analysis of structure-activity relationships and/or the structural analysis of target proteins in the context of ligand screening to define the types of molecules that will most effectively interact with the target protein. Using the invention, key structural features of the target proteins are identified and predictions are made as to the optimal interaction of the target with molecules that may act as antibiotics and inhibitors. Particular target proteins of interest include drug-resistant variants that arise in the course of drug treatment. The invention includes the prediction of such mutations by directed and/or natural evolution, which involves mutation of a target protein, observation of functional changes in the target protein, analysis of the primary and tertiary structural changes responsible for those functional changes and identification of specific chemical structures or molecules which have the appropriate size, shape and chemical properties to modulate the activity of the target protein.

One embodiment of the invention involves a method for identifying a new drug effective against at least one drug-resistant variant of a target protein, comprising: (a) isolating or selecting at least one drug-resistant variant of target protein in the presence of the drug; (b) obtaining the three-dimensional structure of at least one variant or of the drug interaction site of the variant; (c) obtaining

the three-dimensional structure target protein or of the drug interaction site of the target; (d) comparing the three-dimensional structures of the variant(s) and the target protein or of the drug interaction sites of the variant(s) and the target to identify structural similarities and/or differences; and (e) designing or selecting a new drug using the structural similarities and/or differences.

Those of skill in the art will appreciate that the term "selecting" must be read in context. For instance, "selecting" with regard to variants refers to the process of imposing selective conditions, i.e., growing cells that contain the target protein in the presence of the drug, in order to permit the outgrowth, identification and isolation of mutants or variants that are resistant to the selective conditions. In contrast, "selecting" in terms of drug design means, for instance, that a potential inhibitor may be chosen from any existing collection or database and in this regard may be considered to be "selected" rather than "designed."

In order to select or isolate at least one drug-resistant variant of target protein, the organism or cells containing the target protein are exposed to a selecting drug. In general, the selecting drug is a known drug. For example, the selecting drug may be a known drug that has already been observed to give rise to drug-resistant variants or mutations. The target protein is the protein with which the selecting drug interacts.

The drug can be, for example, an antibiotic, an antibiotic resistance inhibitor, a herbicide resistance inhibitor, an insect repellant, an insecticide resistance inhibitor, a viral drug resistance inhibitor, a chemotherapeutic resistance inhibitor, a modulator, an antagonist, an agonist, an effector, a ligand, an antibody, an antibody fragment, a peptide or a small molecule. A modulator is a compound that can interact with the drug-resistant variant(s) of the target protein, thereby rendering drug-resistant variant susceptible to inhibition by the selecting drug or any other drug of interest. In this instance, the drug-resistant variant(s) effectively becomes a target for the selecting drug or a "first generation drug resistant variant." Several "generations" of variants and of new drugs can be selected so that new generations of new drugs may be designed.

The method may include further steps such as contacting a new drug identified by the method with either the drug-resistant variant or the target protein to test the ability of the new drug to interact with the drug-resistant variant(s) and/or the target protein. In this case, "interact" means bind to, *e.g.*, by hydrogen bonding, ionic interactions, van der Waals or hydrophobic interactions, or fit into the structure of the variant or the target protein in such a way so as to form a combined structural entity or a combination thereof, where

the interaction of the new drug with the variant or the target protein preferably results in a change of function, *i.e.*, inhibition of function or an influence on the biophysical properties (i.e. in the case of a modulator). Interaction of a new drug with drug-resistant variants and target proteins may be detected by "assaying" the function of the drug-resistant variants or the target protein in the presence of the new drug, for instance by growing bacteria or viruses or cells expressing the drug resistant variant in the presence of the new drug to determine if the new drug inhibits or decreases multiplication of the bacteria, viruses or cells. Alternatively, interaction of a new drug with drug-resistant variants or the target protein may be shown by co-crystallizing the new drug with the drug-resistant variant(s) or target protein.

The steps of the methods described herein may be repeated to isolate further potential inhibitors or modulators. For instance, a method according to the invention may further include identifying a second generation new drug of at least a second generation drug-resistant variant of the first drug-resistant variant. Such a second generation drug-resistant variant is found using the new drug designed or isolated using the first drug-resistant variant.

Such an extended method might include (a) selecting at least one second generation drug-resistant variant of the first drug-resistant variant by exposing the first drug-resistant variant to the new drug; (b) obtaining the three-dimensional structure of at least one second generation variant or fragment thereof and the three-dimensional structure of the first drug-resistant variant or fragment thereof; (c) comparing the three-dimensional structures of the second generation drug-resistant variant(s) and the first drug-resistant variant or the fragments thereof to identify structural similarities and/or differences; and (d) designing or selecting a second generation new drug using the structural similarities and/or differences.

Other further methodology steps might comprise using SAR (Structure Activity Relationships) and/or another structure/function-based algorithm to design additional potential inhibitors effective against the drug-resistant variant(s) and/or the target protein.

The steps of the method may be repeated in parallel with at least two, or about ten, or more preferably about 50, about 100, about 250, about 500, about 750 or even about 1000 to about 2000 or more variants simultaneously, *i.e.*, in high throughput fashion. Indeed, one of the particular advantages of the present invention is the ability to produce crystals in nano- or microvolumes, permitting simultaneous screening of multiple variants. Tools for performing the methods disclosed herein in high throughput fashion are known and available, for

instance as disclosed in U.S. Patent 6,296,673, which is herein incorporated by reference in its entirety. Any multiwell apparatus would be amenable to use in the parallel screening embodiments disclosed herein, i.e. 24-well plates, 96-well plates, 384-well plates, 1536-well plates, etc.

The invention also encompasses new drugs of drug-resistant variant proteins and native target proteins that are isolated by the methods of the invention, including compositions comprising one or more of such inhibitors. In particular, the invention encompasses any drug, inhibitor or modulator that is targeted against (1) any drug resistant variant (2) a variant containing a novel mutation or (3) a variant containing novel mutation combinations that are isolated or determined by these methods.

The invention encompasses pharmaceutical compositions comprising the new drugs of the invention. Such compositions can include one or more other known drugs to optimally interact with target and variant proteins. Preferred combined therapeutics would include the drug used to select the drug-resistant variant(s) used to design the new drugs of the invention, for instance to be administered in combination with the new drugs of the invention with the goal of preventing, reducing or delaying the development of resistance to the selecting drug.

Drug-Resistant Variants

According to the present invention, drug-resistant variants are proteins that are structurally similar but not identical to the target protein. The structural similarities and differences are specifically generated or identified and structural information relating to those similarities and differences is used to design new drugs.

Drug-resistant variants may be isolated from any available source by one of skill in the art. For example, the drug-resistant variants can be isolated from a patient who is or was undergoing treatment with the drug. A "patient" may be a human or other animal. Drug-resistant variants can be isolated or selected from an agricultural setting. Drug resistance involving the herbicides, insecticides, antibiotics and other drugs used in agriculture is a very big problem. Very often the same antibiotics are used in both humans and animals. For instance, a patient being treated with a particular drug for a bacterial or viral infection or disease may develop a subsequent non-responsive infection mediated by outgrowth of drug-resistant bacteria or viruses or cell population. A variant protein expressed or displayed by such bacteria or viruses or cell population that confers drug-resistance on such bacteria or viruses or cell population would be

considered to be a "drug-resistant variant" according to the present invention. It will be appreciated, however, that the methods could also be performed following the isolation of mutants that exist in the population without performing selection in the presence of the drug, for instance, by detecting mutants using antibodies or by binding to other types of ligands.

Drug-resistant variants may also be selected by at least one cycle of directed evolution. The phrase "at least one cycle" is intended to convey that repetitive directed evolution experiments may be performed on original variants, second generation variants, third generation variants etc. to isolate variants having more than one or two mutations. Directed evolution may be accomplished by any technique known to those skilled in the art, including those selected from the group consisting of random mutagenesis, use of DNA analogs and DNA shuffling. Directed evolution may also be performed using a hypermutator strain of *E. coli*. Such methods are described in more depth in the disclosure to follow.

Preferred drug-resistant variants have at least one structural mutation resulting in a functional change, an activity change or a change in the protein stability as compared to the target protein. Preferred drug-resistant variants may have a combination of two or more mutations resulting in a structural change as compared to the target protein. In addition, drug resistant variants include mutations that confer resistance by stabilizing the overall fold of the protein. Drug-resistant variants would also include variant target proteins expressed by genes having one or more mutations in the signal sequence that lead to increased levels of protein expression, for instance by stabilization of the pre-protein prior to cleavage of the signal sequence.

Drug-resistant variants would also include those with changes in conformation or modification due to the action of accessory proteins, or due to chemical modifications such as phosphorylation/dephosphorylation, sulfation/desulfation, glycosylation/deglycosylation, presence/absence of metal ions/cofactors/substrates/products/ effectors/modulators, or due to changes in folding. Changes in folding can also be due to mutations in chaperone proteins or other proteins rather than the drug-resistant variant itself. In cases where structural changes, stability or instability are achieved through mutation of accessory proteins of the usual drug target, the resulting drug-resistant variants would still be important targets of new drugs based on structural differences, even though they acquire the drug-resistant phenotype indirectly through mutation of a gene encoding another protein. Likewise, such indirect drug-resistant mutants would still be a source of structural data for identifying

domains important for the design of new drugs, inhibitors and/or modulators that interact with either an indirect drug-resistant variants, a direct drug-resistant variant or a target protein.

In this regard, the present invention also encompasses methods for profiling organisms to identify protein targets that are susceptible to conferring drug resistance to the host cells through mutation, including those proteins that do not directly interact with the drug of interest. Such profiling may be accomplished, for instance, by comparing transcripts from drug-resistant cells to transcripts from susceptible cells to identify differences in the level of expression or the sequences of the transcripts expressed. Tools for accomplishing such profiling are disclosed in U.S. Patent 6,333,155, herein incorporated by reference, which discloses a method of comparing gene transcription induced by drug exposure to gene transcription induced by a known mutation to determine if the known mutation is in a candidate target of the drug. The present invention, in contrast, encompasses a method whereby the transcript profiles of drug-resistant and non-resistant cells are compared and used to identify transcripts originating from mutated genes, or transcripts that have increased in abundance due to the activity of mutated genes. Such profiling methodology may be used to identify further targets for the drug design methods disclosed herein.

The drug-resistant variant proteins and mutant proteins identified in the present methods are also an aspect of the invention, as are the nucleic acids encoding them, including the genes, cDNA, mRNA, genomic nucleic acids and nucleic acid fragments thereof. Structurally homologous variants of these variants, *i.e.*, genes encoding proteins containing mutations that do not change the amino acid sequence or lead to conservative substitutions that do not alter the structure of the variant, are also included, as are libraries comprising the variants, proteins and mutant proteins disclosed herein, or cells expressing the same. Libraries of mutant proteins may be used to test candidate drugs as disclosed in U.S. Patent 6,063,562, which is herein incorporated by reference. Alternatively, such libraries could be used in the parallel screening or high throughput drug design methodology disclosed herein.

Variant proteins and second-generation or further generation variant proteins will find other uses than in the inhibitor design and screening methods disclosed herein. For instance, such variants could form the basis of vaccine compositions for administration to either humans or animals, for the purpose of raising a protective immune response against drug-resistant pathogens expressing or displaying variant proteins. Such drug-resistant variants may also

be used to raise antibodies (both for diagnostic and/or neutralizing purposes) and antisera using methods known in the art or may be used to identify other non-inhibitory ligands, which could then form the basis of diagnostic assays to be used to screen patients for the presence of drug-resistant cells, tumors or pathogens. Kits based on such assays, for instance which include the drug-resistant variant proteins described herein and/or antibodies or ligands specific thereto, are also envisioned.

Drugs

A selecting drug as well as a new drug of the invention can an antibiotic, an enzyme inhibitor, a protein inhibitor, an antibiotic resistance inhibitor, a herbicide resistance inhibitor, an insect repellant, an insecticide resistance inhibitor, a viral drug resistance inhibitor, a chemotherapeutic resistance inhibitor, a modulator, an antagonist, an agonist, an effector, a ligand, an antibody, an antibody fragment, a peptide or a small molecule.

New drugs of the invention may be selected for the ability to interact with a drug-resistant variant(s) and/or the target protein. In some instances the new drug interacts with structurally distinct regions of a variant protein that may not be present in the native target protein. Alternatively, a new drug, e.g. a potential inhibitor or modulator, may be selected for its ability to interact with regions of a drug-resistant variant(s) that are structurally similar to those in the native target protein or to bind to regions that are structurally critical to both target and variants.

A target protein "region" or "interaction site" according to the present invention can be an active site, allosteric site, regulatory site, global suppressor site, as well as any mutation site located in any region of the sequence or structure of a target protein that influences any biophysical property, i.e., activity, stability, expression level, or biological function, of these aforementioned regions. Such regions would include, but would not be limited to, domains and domain interfaces of a protein, and parts or fragments thereof.

A "domain" may be defined structurally as a compactly folded part of a protein that is linked to other domains by one or few other structural elements such as a loop or helix. A domain may also be defined dynamically as a

relatively rigid region in a protein that is separated from other domains by more flexible interdomain regions. A dynamical domain may be isolated or identified by calculating the large scale movements of certain points over the entire protein, either through normal mode analysis or by comparing two or more structures of the same protein that have domain movements. The parts of a protein that move in synchronous fashion are grouped in dynamical domains, with a domain "interface" being composed of regions in a domain that interact with regions in an adjacent domain. Such interactions are particularly important near the hinge regions between domains. The present invention allows for the identification of protein regions and domains and domain interfaces that are important for the development of drug resistance; therefore, once such regions are identified, a potential inhibitor may also be selected for its ability to inhibit the folding of one or more important regions or domains of the drug-resistant variant or the target protein.

New drugs may be selected from a database of inhibitors or ligands, for instance using computer modeling and/or high throughput screening. Preferred databases for use in the present invention include databases of small organic or inorganic molecules. Libraries of small compounds which find use herein will generally comprise at least 2 compounds, often at least about 25 different compounds, more often at least about 100 different compounds, and usually at least about 200 to 300 different compounds. Such a library of organic compounds to be screened against the modified polypeptide of interest may be obtained in a variety of ways including, for example, through commercial and non-commercial sources, by synthesizing such compounds using standard chemical synthesis technology or combinatorial synthesis technology (see Gallop *et al.*, J. Med. Chem. 37:1233-1251 (1994), Gordon *et al.*, J. Med. Chem. 37:1385-1401 (1994), Czarnik and Ellman, Acc. Chem. Res. 29:112-170 (1996), Thompson and Ellman, Chem. Rev. 96:555-600 (1996), and Balkenhohl *et al.*, Angew. Chem. Int. Ed. 35:2288-2337 (1996)) and by obtaining such compounds as degradation products from larger precursor compounds, e.g. known therapeutic drugs, large chemical molecules. A large number of small chemical compounds are readily obtainable from commercial suppliers such as Aldrich Chemical Co., Milwaukee, Wis. and Sigma Chemical Co., St. Louis, Mo.

It should be understood that, although the phrase "new drug" is written as a singular entity, a new drug may comprise a combination of two or more individual entities, *i.e.*, at least two peptides or small molecules, where the two entities interact in combination or individually with the variant or target protein. New drugs may be selected using drug design with any one or more of the methods selected from the group consisting of structure based drug design (SBDD), virtual ligand screening, ligand screening, high-throughput screening and ultra high-throughput screening as discussed further in depth below. For instance, U.S. Patent 6,297,021 discloses methods for structure-based drug design and is herein incorporated by reference in its entirety.

Preferred new drugs of the invention are inhibitors and modulators of the variant protein. Preferred new drugs of the invention are small molecules that generally are not large proteins or polynucleic acids. Compounds that find use in the present invention include, for example, aldehydes, ketones, oximes such as O-alkyl oximes, preferably O-methyl oximes, hydrazones, semicarbazones, carbazides, primary amines, secondary amines such as N-methylamines, tertiary amines such as N,N-dimethylamines, N-substituted hydrazines, hydrazides, alcohols, ethers, thiols, thioethers, thioesters, disulfides, carboxylic acids, esters, amides, ureas, carbamates, carbonates, ketals, thioketals, acetals, thioacetals, aryl halides, aryl sulfonates, alkyl halides, alkyl sulfonates, aromatic compounds, heterocyclic compounds, anilines, alkenes, alkynes, diols, amino alcohols, oxazolidines, oxazolines, thiazolidines, thiazolines, enamines, sulfonamides, epoxides, aziridines, isocyanates, sulfonyl chlorides, diazo compounds, acid chlorides, and the like.

Virtually any molecule that is capable of bonding covalently or non-covalently or by any other means to a drug-resistant variant and/or target protein as disclosed herein may find use in the present invention.

Suitable Target Proteins

Suitable target proteins for use in the disclosed methods include proteins selected from the group consisting of bacterial proteins, viral proteins, microbial proteins, fungal proteins, cellular proteins contributing to the development of cancer, cellular proteins involved in inflammation or autoimmune disorders or

diseases, and agricultural targets, *e.g.* herbicide resistance, insect resistance, etc. A suitable target protein of the present invention is any protein that is the target of a drug, *i.e.*, antibiotic, chemotherapeutic, antibody reagent, herbicide, insecticide or any other inhibitor, and which may undergo structural changes either by mutation or other means to confer drug-resistance to the pathogen or cell in which it is expressed.

A target protein may be found in any pathogenic organism, but preferred target organisms include those selected from the group consisting of *Haemophilus influenzae, Staphylococcus aureus, Streptococcus pneumoniae, Enterococcus faecalis, Escherichia coli, Salmonella, Pseudomonas aeruginosa, Bacillus anthracis, Yersiniia pestis, Francisella tularensis, Brucella, Coxiella burnetii, staphylococci, Clostridium botulinum, Bordetella pertussis, Neisseria meningitidis* Z2491, *Streptococcus mutan, Klebsiella pneumoniae, Chlamydia pneumoniae, Chlamydia trachomatis, Heliobacter pylori, Mycoplasma pneumoniae, Legionella pneumophila, Legionella micdadei, Treponema pallidum, Toxoplasma gondii, Borrelia burgdorferi, Campylobacter jejuni, Listeria monocytogenes, Neisseria gonnorhea, Neisseria meningitidis, Rickettsia prowazekii, Mycoplasma genitalium, Mycobacterium tuberculosis, Streptococcus pyogenes, Vibrio cholerae, Trypanosoma cruzi, Cryptosporidium parvum,* human immunodeficiency virus (HIV), cytomegalovirus (CMV), hepatitis C virus (HCV), herpes simplex virus (HSV), bovine diarrhea virus, human rhinovirus (HRV), hepatitis B virus (HBV), influenza virus, variola virus (smallpox), viral encephalitides, arenaviridae virus, bunyaviridae virus, filoviridae virus, flaviviridae virus, and filamentous fungi (Fusarium, Myrotecium, Cephalosporium, Trichoderma, Verticimonosporium, Stachybotrys species).

Potential cancer targets include cancer of the breast, prostate, throat, ovary, lung, brain, pancreas, stomach, and skin to name a few, as well as cancers of the blood including all forms of leukemia, lymphoma, myeloma, etc.

Potential immune targets include proteins in cells involved in inflammatory and autoimmune diseases. Examples of immune, allergic and inflammatory disorders or diseases that may be targets of the present invention include general inflammation, skeletal-muscular disorders, osteoarthritis, gout, asthma, lung edema, adult respiratory distress syndrome, pain, aggregation of platelets, shock, rheumatoid arthritis, juvenile rheumatoid arthritis, psoriatic arthritis, psoriasis, autoimmune uveitis, allergic encephalomyelitis, systemic lupus erythematosis, acute necrotizing hemorrhagic encephalopathy, idiopathic thrombocytopenia,

polychondritis, chronic active hepatitis, idiopathic sprue, Crohn's disease, Graves ophthalmopathy, primary biliary cirrhosis, uveitis posterior, interstitial lung fibrosis; allergic asthma; and inappropriate allergic responses to environmental stimuli such as poison ivy, pollen, insect stings and certain foods, including atopic dermatitis and contact dermatitis.

Preferred target proteins may be selected from the group consisting of enzymes, transcription factors, DNA polymerases, RNA polymerases, ribosomal proteins and DNA binding proteins, or any protein which constitutes or comprises a site for antibiotic or inhibitor action or a site for antibiotic resistance. For instance, target proteins may be selected from the group consisting of Topo IV B subunit (ParE), gyrase B subunit (GyrB), β-lactamases, penicillinases, cephalosporinases, choramphenicol acetyltransferases, aminoglycoside modifying enzymes, D-ala D-ala ligase (ddl), Peptide deformylase, FabI, GlmU, MurA, MurB, MurC, MurD, FtsZ, Dam methylase, emr/mmr multidrug resistance, bacterial DNA methylase, squalene synthase, human immunodeficiency virus (HIV) reverse transcriptase, HIV protease, HIV integrase, HIV TAT/TAR, HIV NCp7, feline immunodeficiency virus protease, CCR5/CXCR4, cytomegalovirus (CMV) polymerase, HCV NS3 helicase, HCV NS3-4A (Ser protease), HSV protease, HSV TS, Bovine diarrhea virus NS5B (RNA synthesis), HRV 3C protease, HBV PPK (polyphosphate kinase), HMG Co-A reductase, Chitin synthase-1, D- glucan synthase, inositol phosphorylceramide synthase (IPC), EF-2, neuraminidase, topoisomerase, thrombin, Factor Xa, cathepsins B, C, S and K, caspases, matrix metalloproteinases, Cyclin-dependent kinases, urokinase, EGFR tyrosine kinase, Janus kinase family (JAK), STAT, Bcl-2, receptor family tyrosine kinase, insulin receptor kinase, protein-tyrosine phosphatases, Src family kinases, Lck family kinases, ZAP/SYK family kinases, and TEC family tyrosine kinases, angiotensin-converting enzyme, monoamine oxidase and tetrahydrofolate reductase. A particularly preferred target protein is beta-lactamase, with the selecting drug being any penicillin derivative or beta-lactam antibiotic, including cefotaxime as described in the non-limiting examples included *infra*.

Antibiotic Resistance Target Proteins

As described above and used herein, target proteins include those proteins which constitute a site for drug, antibiotic or inhibitor action, or a site for drug or antibiotic resistance. Any protein known by one of skill in the art to interact with an antibiotic, whose three-dimensional structure and function can be determined, is contemplated. The present methods therefore permit

development of new drugs such as new antibiotics, new inhibitors, new modulators and new compounds to combat antibiotic resistance which can draw upon any information available to one of skill in the art about these antibiotic, inhibitor, and antibiotic-resistance mechanisms.

As illustrated in the following table, many antibiotics have well characterized targets of action. Any of these targets provides a suitable target protein for use in the methods of the present invention:

Antibiotic Family	Example	Target
β-lactam antibiotics	penicillins, cephalosporins	cell wall biosynthesis
Sulfonamides	Sulfanilamide	blocks synthesis of tetrahydrofolate
Aminoglycosides	streptomycin	protein synthesis
trimethoprim	--	folate metabolism
Chloramphenicol	--	protein synthesis
Vancomycin	--	cell wall synthesis

Mechanisms of antibiotic action are further described in the table below.

Antibiotic Type	Type of Action	Effect
Psicofuranine	Inhibits xanthylic acid aminase	Prevents synthesis of GMP, inhibits DNA synthesis
Nalidixic acid	Replicator	Blocks DNA synthesis
Griseofulvin	Replicator	Blocks DNA synthesis
Mitomycin	Binds complementary strands of DNA	Inhibits DNA replication by inhibiting strand separation
Actinomycin D	Binds to guanine in DNA	Inhibits RNA synthesis
Rifampicin	Binds to RNA polymerase	Prevents RNA synthesis
Rifamycin	Binds to RNA polymerase	Prevents RNA synthesis
Streptomycin Neomycin Kanamycin	Binds to 30S ribosomal subunit	Inhibits protein synthesis by stimulating miscoding and inhibiting peptide chain elongation
Tetracycline Edeine	Binds to 30S ribosomal subunit	Inhibits binding of amino acyl-tRNA to ribosome
Puromycin	Binds to 50S ribosomal subunit	Causes premature release of incomplete polypeptide chains
Chloramphenicol	Binds to 50S ribosomal subunit	Inhibits attachment of amino acyl-tRNA to ribosome
Erythromycin	Binds to 50S ribosomal subunit	Inhibits protein synthesis
Lincomycin	Binds to 50S ribosomal subunit	Prevents peptide bond formation

Antibiotic Type	Type of Action	Effect
D-cycloserine (oxamycin)	Inhibits racemase transforming L to D alanine	Prevents synthesis of peptide side chain on muramic acid
Vancomycin Risocetin Bacitracin	Inhibits growth of glycopeptide polymer	Prevents synthesis of murein
Penicillin Methicillin Ampicillin Cephalosporin	Inhibits transpeptidation of side chain to bridge	Prevents crosslinking of chains of glycopeptide
Lysostaphin	Hydrolyzes peptide side chain and cleaves muramic acid-glucosamine polymer	Lysis of staphylococci
Tyrocidine	Damage to membrane; decreases respiration	Leakage of cell contents
Gramicidin	Uncouples oxidative phosphorylation; binds to membrane	Leakage of cell contents
Polymycin Colistin	Releases proteins from membrane	Leakage of cell contents

Antibiotic resistance is commonplace. Many mechanisms for antibiotic resistance are known, including over-expression of target enzymes, expression of antibiotic inactivating enzymes, and mutation of a target protein so that it is no longer recognized by the antibiotic. Preferred target proteins of the present invention undergo structural changes that confer antibiotic resistance, for instance because of internal mutations or the action of other proteins. Examples of some mechanisms of antibiotic resistance are given below:

Antibiotic	Principle mechanism of resistance
Penicillins and other β-lactam antibiotics	Inactivation by β-lactamase
Sulfanilamide	Mutation of dihydropteroate synthase
Aminoglycosides	Inactivation by aminoglycoside modifying enzyme, or by target mutation
Trimethoprim	Mutation of dihydrofolate reductase
Chloramphenicol	Inactivation by chloramphenicol transacetylase
Methicillin	Mutation of penicillin binding proteins
Vancomycin	Mutation in target cell wall peptide

Inactivation of an antibiotic is probably the most common mechanism for drug resistance. For example, the most common form of resistance to β-lactam antibiotics is the production of β-lactamases, which degrade the antibiotic molecule. Other enzymes involved in drug resistance include penicillinases, cephalosporinases, choramphenicol acetyltransferases and aminoglycoside modifying enzymes. Such enzymes inactivate antibiotics by modifying them to inactive compounds. Other mechanisms which contribute to antibiotic resistance include drug permeability mutations, expression of transport proteins that actively extrude antibiotics from target organisms, and mutations in the cellular targets of antibiotics.

Recently, two studies surveyed the genetic diversity present in isolates obtained from single hosts harboring the gastric pathogen *Heliobacter pylori* (D.A. Israel *et al.*, Proc. Natl. Acad. Sci. USA 98:14625-30 (2001); B. Björkholm *et al.*, Infect. Immun. 69:7832-8 (2001)). In both studies, multiple genetic variants of *H. pylori* were found, with some isolates having new genetic material, while other strains taken from the same host had missing genes. Thus, bacteria are in a constant state of genetic flux, especially when antibacterials are administered or during other potentially lethal challenges, such as survival in a host organism. Since new forms of many genes generally occur in all organisms (especially pathogenic microbes and viruses, as well as neoplastic cells) the approach outlined in the present invention would also be applicable to any such newly acquired or newly expressed genes that are structurally modified to convey drug resistance.

Anti-Viral Drugs

The methods of the present invention are also applicable to the design of new drugs for the treatment of viral infections, and may be applied to any virus that has been isolated. Preferably, the virus is characterized by one of skill in the art with respect to gene expression, protein content and structure. In this embodiment, any viral protein that undergoes structural changes to confer drug resistance, or any cellular protein that undergoes structural changes to confer drug resistance to an infecting virus, may be a target protein for use in the disclosed methods. Such methods find particular utility in light of the fact that a reasonable estimate is that a significant number of infectious diseases found in developing countries are caused by viruses (J.N. Delgado & W.A. Remers, Eds "Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical

Chemistry," Tenth Edition, Lippincott-Raven, NY (1998)), as well as by bacterial infections.

The viruses targeted by the present invention may be either DNA or RNA type viruses. Both types are obligate cellular parasites that depend upon a host for energy and the biochemical substances necessary for replication and protein synthesis. Thus, most existing drugs target virus-specific enzymatic processes required for the viral life cycle. Any existing drug or future drug may be used in the methods disclosed herein in order to predict future variants showing resistance to the drug. Two notable examples of successful antivirals developed with structure-based drug design include the influenza virus neuraminidase inhibitors targeted to block the release of mature virus particles from host cell sialic acid residues, and the HIV protease and HIV reverse transcriptase enzyme inhibitors.

Structure-based drug design was used to develop the neuraminidase inhibitors oseltamivir and zanamivir (Tamiflu and Relenza) (Y.S. Babu *et al.*, J. Med. Chem. 43:3482-6 (2000)). Both of these drugs have been shown to be clinically effective in the prophylaxis and treatment of influenza A and B, and have demonstrated the development of minimal viral resistance in initial clinical studies (I.R. McNicholl *et al.*, Ann. Pharmacother. 35(1):57-70 (2001); M.G. Ison *et al.*, Curr. Opin. Pharmacol. 1(5):482-90 (2001); N.A. Roberts Prog. Drug Res. Spec No:35-77 (2001)). Although inhibitor-resistant forms of neuraminidase have been generated using *in vitro* passage in the presence of the drugs oseltamivir and zanamivir (M.T. Murrell *et al.*, J. Virol. 75(14):6310-20 (2001); L.V. Gubareva *et al.*, Antiviral Res. 53(1):47-61 (2002)), mutant forms of neuraminidase have started to appear in clinically-derived isolates (L.V. Gubareva *et al.*, J. Infect. Dis. 183(4):523-31 (2001)). Thus, new inhibitors specific for influenza neuraminidase are needed, and the methods of the present invention would find particular utility in analyzing the structure of existing and future drug-resistant variants so that new inhibitors may be designed.

HIV protease and HIV reverse transcriptase have both been used as targets for structure-based drug design against HIV-1 retroviruses, although HIV-1 protease has been the most important target in the development of antiviral therapies against HIV-1 infection (Todd *et al*, Proteins 36:147-156 (1999)). The crystal structures of many wild type and mutant HIV proteases with and without bound inhibitors have been determined. This information has been very valuable and has led to development of the structure-based inhibitor Nelfinavir (also known as Viracept) (Kaldor *et al*, J. Med. Chem. 40:3979-85 (1997)). In addition, many experiments have determined which HIV protease

mutations are selected for during inhibitor monotherapy and combination therapy. Table 1 lists the mutations that are observed during protease inhibitor monotherapy, and there is a database available that lists sequence information for a large number of published HIV reverse transcriptase and HIV protease isolates, including inhibitor resistance information (http://hivdb.stanford.edu/).

The methods of the present invention can be employed to analyze the three-dimensional structures of these isolates, and to design new inhibitors that target important domains of these enzymes with an understanding of how the proteins conform structurally to avoid the action of the drug.

DRUG RESISTANT HIV PROTEASE MUTATIONS

SAQUINAVIR RESISTANT MUTATIONS

Mutation(1)	Location(2)
G48V	flap region
L90M	adjacent to active site
L10I	distant from active site
L63P	distant from active site
I84V	active site
M46L	flap hinge region
*A71V	distant from active site
M36I	distant from active site
* putative global suppressor	

NELFINAVIR RESISTANT MUTATIONS

Mutation(2, 3, 4)	Location(2)
D30N	active site
*A71V	distant from active site
I84V	active site
M46I	flap hinge region
* putative global suppressor	

AMPRENAVIR RESISTANT MUTATIONS

Mutation(2, 5)	Location(2)
I84V	active site
L10F	distant from active site
I50V	flap region
M46I	flap hinge region
I47V	flap hinge region
D60V	distant from active site

RITONAVIR RESISTANT MUTATIONS

Mutation (2, 4,7)	Location(2)
V82A	active site
I54V	Flap
*A71V	distant from active site
M46I	Flap
K20R	distant from active site
* putative global suppressor	

(1) Jacobsen *et al.*, J. Infect. Dis. 173:1379-87 (1996)
(2) Boden *et al.*, Antimicrob. Agents & Chemother. 42(11):2775-83 (1998)
(3) Patick *et al.*, Antimicrob. Agents & Chemother. 40(2):292-7 (1996)
(4) Mahalingam *et al.*, Eur. J. Biochemm. 263:238-45 (1999)
(5) Partaledis *et al.*, J. Virol. 69(9):5228-35 (1995)
(6) Molla *et al.*, Nature Medicine 2(7):760-6 (1996)
(7) Markowitz *et al.*, J. Virol. 69(2):701-6 (1995)

Under native conditions, the dimeric HIV-protease molecule does not exist in a single conformation (Todd *et al*, 1999). The flap region located at the top of the protease binding site has very little structural stability, leading to many conformational states in the absence of substrate or inhibitor. The bottom of the binding site is the most stable part of the protein and contains the catalytic active site residues (residues 24-26). Upon drug, substrate or inhibitor binding, the flap region takes on a more ordered conformation. Todd *et al*, have shown that the effects of ligand binding and stabilization of the flap region are propagated throughout the molecule and most significantly to the bottom of the binding site (containing the catalytic residues). In contrast, the effects due to the interactions with the sides and the bottom of the active site are more local.

Resistance to the clinically important inhibitors saquinavir, nelfinavir, amprenavir, and ritonavir have shown the development of resistance mutations in three "hotspot" areas (Table 1): residues M46, I47, G48, G49, and I50 are located in the flap region, a second group consists of residues R8, D25, G27, A28, D29, and D30 located in the bottom of the molecule near the active site, and the third group consists of residues P81, V82, and I84 located on the sides of the active site (Todd *et al*, 1999). As expected, mutations at residues located in these hotspot regions have substantial effects on inhibitor resistance (Boden *et al*, 1998). Mutations located in the active site or within the flap region reduce inhibitor binding significantly but lead to protein destabilization and reduced viral reproduction (Jacobsen *et al*, 1996)(Boden *et al*, 1998). Compensatory mutations located outside of the binding site region stabilize the protease and improve viral reproductive capability (Markowitz *et al*, 1995). Upon inhibitor challenge above certain thresholds, active site mutations have been shown to revert to wild type and are replaced with more efficient active site or flap region mutations (Mahalingam *et al*, 1999) (Partaledis *et al*, 1995). Thus, the

development of therapeutics containing a combination of different inhibitors is an important goal of treatment.

The hotspot binding regions are structurally separated in the two HIV protease domains, and the two domains are linked by four strands. The first domain consists of a three β-strand sheet containing the flap region. The second domain contains the side and bottom regions of the active site, as well as the remainder of the protease molecule. Four strands link the two domains: residues 32-42 (11 residues), residues 61-64 (4 residues), residues 66-73 (8 residues), and residues 78-83 (6 residues). The mutation A71V is located on strand 66-73 which connects the end of a β-sheet in the first domain (flap region) with the end of a β-sheet in the second domain.

In spite of the many resistance mutations that arise throughout the HIV protease molecule, any mutational effects that initiated from the first domain must still propagate to the second. In the position occupied by the A71V mutation, the two domains are tethered together, preventing gross distortions between the two domains. Biochemical studies support this hypothesis of A71V stabilization. In addition, the A71V mutation is selected for in nelfinavir, ritonavir, as well as saquinovir resistance. Two additional mutations, one at position 84 in the active site region and one at position 46 in the flap region, are selected for in response to three or more inhibitors (Table 1). The A71V mutation is the only mutation not located in the active site or flap region mutation that is also selected for in response to three or more different inhibitors, emphasizing its importance. Furthermore, the A71V mutation is capable of reversing destabilizing mutations at multiple locations throughout the enzyme (Jacobsen *et al* (1996), Boden *et al* (1998), Patick *et al* (1996), and Markowitz *et al* (1995)), consistent with a possible role as a global suppressor mutation. Thus, a preferred HIV inhibitor of the present invention will be designed with consideration of the A71V mutation.

Utilization of a protease knockout of HIV-1 virus will enable the in-vitro evolution of the HIV-protease gene using methods of directed evolution (e.g. gene-shuffling, error-prone PCR). After rounds of evolution have been completed the protease gene will then be transfected back into the knockout virus in preparation for in-vivo selection.

In order to carry out the in-vivo selection human lymphoblastoid T-cell lines such as CEM-SS or MT-4 are used to propagate the HIV-1 virus transfected with evolved protease gene. The virus will be passaged in the selected lymphoblastoid T-cell line in the presence of a previously established HIV-protease inhibitor. After suitable adsorbtion period the cells are pelleted, washed, resuspended and plated on RPMI media plates containing a predetermined concentration of protease inhibitor. After a certain period of time lymphocyte colonies with no cytotoxic effect or syncytium formation were picked and cultured separately. The cell-free viral supernatant is collected and the protease gene is amplified from cellular DNA extracts by PCR and subjected to further rounds of evolution and selection.

After the desired level of protease resistance is selected for the resulting protease genes are sequenced and cloned into an expression vector for in-vitro analysis and structure determination.

In order to assay the activity of an evolved or drug-resistant HIV-protease in the presence of increasing concentrations of protease inhibitor the protease is combined with a solution buffered at an appropriate pH possibly containing dithiothreitol and EDTA. This stock solution is added to vials containing a predetermined concentration of protease inhibitor and allowed to incubate for a period of time. The reaction is initiated by the addition of a p-nitrophenylalanine substrate and incubated. After incubation the reaction is quenched with trifluoroacetic acid and the products quantified with HPLC or mass spectrometry. The data can be fitted to a competitive inhibition model and the Ki and IC50 determined. See Partaledis, J.A., et. al. In Vitro selection and characterization of human immunodeficiency virus type 1(HIV-1) Isolates with reduced sensitivity to hydoxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J. Virol.* **69(9),** 1995 5228-5235.

Anti-Cancer Agents

Another important use of the methods described in this application involves the development of new small molecule anticancer agents that will be less amenable to the development of inhibitor resistant escapes when administered to a neoplastic cell population. A number of different signal

transduction pathways become deregulated in human cancers, leading to uncontrolled cellular proliferation. A number of different anticancer therapies exist; selected anticancer agents where resistance mechanisms have been studied are included in the following table.

5

Anticancer Agents

Anticancer Agent	Target & Mechanism	Mechanism(s) of Resistance
Methotrexate (Rheumatrex®)	Antimetabolite (folic acid analog) that inhibits Dihydrofolate Reducatase (DHFR), butalso directly inhibitors folate-dependent enzymes of de novo purine and thymidylate synthesis pathways. Drug targets include: Dihydrofolate Reductase Thymidylate synthase Glycinamide ribonucleotide transformylase	Mechanisms of resistance include: Impaired transport of methotrexate into cells Production of altered forms of Dihydrofolate reductase (DHFR) that have decreased affinity for inhibitor Increased concentrations of intracellular DHFR Decreased ability to synthesize methotrexate polyglutamates Decreased thymidylate synthase activity
Cytarabine (Cytosine Arabinoside) [Cytosar-U®]	Antimetabolite (pyrimidine analog) that inhibits DNA synthesis Drug targets include: Deoxycytidine kinase Cytidine deaminase DNA Polymerase	Resistance is poorly understood, but mechanisms of resistance include: Deficiency of deoxycytidne kinase Increased Cytidine Deaminase activity Reduced affinity for DNA polymerase
Dacarbazine (DTIC) [Dtic-Dome®]	Alkylating agents that target DNA synthesis enzymes including: Guanine-alkyl transferase (DNA repair enzyme) Glutathione transferase	Resistance is poorly understood, but mechanisms of resistance include: Decreased permeation of actively transported drugs Increased production of nucleophilic substances, such as glutathione that can compete with the target DNA for alkylation Increased activity of the DNA repair enzymes
Imatirib Mesylate (Gleevec™)	Drug target: Bcr-Abl Kinase	Resistance caused by single point mutation, leading to reactivation of kinase activity.
Doxurubicin (Adriamycin,	Anthracylcine antibiotic that targets DNA and	

Anticancer Agent	Target & Mechanism	Mechanism(s) of Resistance
Rubrex) Mitoxantrone	RNA synthesis Topoisomerase II Cytochrome P450 reductase	

Recently, the successful small molecule drug candidate Gleevec was made using rational drug design, and has been shown to be an effective *in vivo* inhibitor of the Bcr-Abl tyrosine kinase responsible for the genetic malfunction present in chronic myelogenous leukemia (CML) (B.J. Druker *et al.*, Blood (Suppl.) 94:368a (1999);T. Schindler *et al.*, Science 289:1938-42 (2000)). Unfortunately, as is the situation with most conventional anticancer drugs, patients with advanced forms of cancer often relapse, with their tumor cells becoming resistant to the chemotherapeutics. In the case of Gleevec, a single point mutation in the Bcr-Abl kinase (Thr315 Ile) was found in all relapsed patients, and was found to cause reactivation of the kinase activity. The methods as outlined herein can be used to predict new sites of mutation, as well as new mutant combinations that can arise upon application of inhibitor selection during a directed evolution procedure. The new Bcr-Abl kinase mutant proteins can then be used as targets for design of new inhibitors based on SBDD, VLS, or any of the ligand screening procedures applied to the new mutants or both the new mutant and wild-type protein targets.

Inhibitors of topoisomerases are widely used in the treatment of cancer, including inhibitors of topoisomerase I (camptothecin analogs such as irinotecan and topotecan) and topoisomerase II (etoposide and doxorubicin). Topoisomerase drug resistance is commonly due to overexpressing P-glycoprotein or multidrug resistance-associated protein, however, some camptothecin (CPT)-resistant topoisomerase I (top1) proteins have been isolated, harboring specific top1 mutations that are related to their CPT-resistance. For example, the R364H mutation is close to the catalytic tyrosine and other well-known mutations leading to CPT resistance (Urasaki et al. Cancer Res 61(5):1964-9 (2001). In addition, CPT-resistant cell lines with their corresponding mutations have been isolated for human prostate carcinoma cells DU-145/RC1 (mutation R364H), Chinese hamster fibroblasts DC3F/C10 (mutation G503S), and human leukemia CEM/C2 cells (N722S) Urasaki et al. Cancer Res 61(2):504-8 (2001). Structure-based drug design has recently been proposed for wild-type and mutant forms of topoisomerase (http://dtp.nci.nih.gov/branches/gcob/gcob_web25.html). However, the methods as outlined herein can be used to design new inhibitors of both the wild-type and

mutant topoisomerase proteins based on SBDD, VLS, or any of the ligand screening procedures applied to both protein targets.

Some cancers result from the mutation of genes involved in the control of cell division, for example, tumor suppressor genes such as p53. One embodiment of the present invention would therefore involve the design of drugs that restore activity to proteins that normally act to control cell division, including p53, but that have mutated, thereby causing or contributing to malignant transformation or remission or recurrence of malignancy following drug treatment. A preferred novel drug according to the present invention would stabilize all mutant forms of p53, for instance, that would be predicted to arise during the course of a drug treatment. Such drugs could be administered individually or as part of a combined therapeutic regimen.

Directed Evolution Methods

As used herein, "directed evolution" refers to the analysis of target protein structure and function by mutagenesis, phenotypic selection, screening or assessment, function evaluation, and/or sequence analysis. A "directed evolution profile" for a target protein is a collection of structural data comprising the structural changes that have occurred as the target protein evolves into a series of drug-resistant variants. Other data can be included in the directed evolution profile for a target protein such as protein activity data, drug-resistance data, protein stability data, frequency of mutation and the like.

Several references disclosing directed evolution techniques are discussed below and are incorporated by reference in their entirety at the location of these citations. The molecular cloning techniques required for an underlying appreciation of directed evolution methodology are now commonplace, as those of skill in the art will appreciate. For a general review of cloning and expressing genes and analyzing mutations, reference may be made to Sambrook *et al.*, Molecular Cloning, A Laboratory Manual 2nd ed., Cold Spring Harbor Laboratory Press, 1989 (which is herein incorporated by reference for all it teaches and all the methods it references).

Natural selection does not try out all possible combinations of mutations at random, but rather proceeds in a series of steps, each step leading to a phenotype that is more fit than the previous generation. The appearance of mutations in antibiotic resistance genes, viral genes, or cancer-related genes is selected for by the use of antibiotics, inhibitors, or other therapeutics in clinical settings. For example, natural evolution of β-lactamase has taken place in response to an array of different penicillin derivatives, cephamycins, and four

generations of cephalosporins. Even though there are a large number of different antibiotics prescribed in the clinic, only 13 sites of mutation (out of 290 total residues) in the β-lactamase gene are sufficient to give rise to extended-spectrum β-lactamase antibiotic resistance.

Directed evolution is a powerful tool to quickly identify key structural changes that can give rise to antibiotic resistance. As provided by the present invention, directed evolution experiments have determined mutation hotspots, have correctly predicted the clinical emergence of a three-mutant combination, and additionally have predicted the clinical emergence of even higher order combinations of these mutations. Structural analysis complements directed evolution in the methods of the invention. For instance, the present invention describes the structural basis of increased cefotaxime resistance due to the E104K/M182T/G238S combination of mutations.

In general, mutations isolated early in the directed evolution process often act by reorganizing the active site topology, while later mutations can act, for example, as compensatory mutations. For example, the later-isolated M182T mutation is a global suppressor mutation that is located in the hinge region between the two domains of β-lactamase. The M182T mutation may act by stabilizing the proper orientation of the active site residues. Therefore, according to the present invention, a new class of antibiotics or inhibitors can be designed to inhibit folding of the native β-lactamase enzyme conformation.

Any method available to one of skill in the art can be used for generating mutations to perform directed evolution. Methods contemplated by the present invention include, for example, the use of natural selection, DNA analogs, DNA shuffling, hypermutator *E. coli* strains, and random mutagenesis methods. These methods can vary in the quantity and type of mutations that they produce, but all have the potential to mimic natural evolution. Accordingly, combinations of mutagenesis methods are also contemplated.

Directed evolution methods can introduce mutations at rates of up to about 3 percent (Stemmer, 370 Nature 389-91 (1994); Zaccolo *et al.*, 285 J. Mol. Biol. 775-83 (1999)). A mutation rate of about 0.1% is equivalent to about one amino acid mutation per one kilobase DNA. The directed evolution method using the incorporation of DNA analogs to introduce mutations has been found to mutate particular bases more frequently than others (Zaccolo *et al.* (1999)), resulting in nonnatural and nonrandom distributions (Skandalis *et al.*, 4 Chem. Biol. 889-98 (1997)). DNA shuffling (Stemmer (1994)) applies a low mutagenesis rate using polymerase error, and additionally provides for DNA

recombination, which is an efficient method to create new combinations of drug resistance mutations.

Methods for performing mutagenesis by the DNA analog method are described in Zaccolo *et al.* (1999). Mutagenesis by the *in vitro* DNA shuffling method can be performed as described in Stemmer (1994), which is herein incorporated by reference. In one embodiment, the *in vivo* DNA shuffling strategy is used that takes advantage of mutagenesis using a hypermutator strain of *E. coli* by the methods described in Echols *et al.*, 80 Proc. Nat'l. Acad. Sci. USA, 2189-92 (1983) and Low *et al.*, 260 J. Mol. Biol. 359-68 (1996). This procedure removes the PCR step necessary with *in vitro* DNA shuffling and the DNA analogs method.

Prior to the first directed evolution study, there were nineteen naturally occurring TEM isolates (Stemmer (1994)). In the year 2000, there were over ninety TEM β-lactamase mutants (Jacoby *et al.*, http://www.lahey.org/studies/webt.htm (2000)). The DNA analog and DNA shuffling methods have each identified mutations at five positions in wild type β-lactamase (TEM-1) that have also been obtained in clinical isolates (Fig.1). DNA shuffling also predicted three mutations at positions 42, 92 and 182 before their clinical isolation (Fig. 2). In a 1994 study, DNA shuffling predicted the E104K/M182T/G238S combination of mutations to be added to the TEM-1 isolate, leading to a mutant with an observed 500-fold increase in antibiotic resistance relative to wild type TEM-1 (Stemmer (1994)). Four years later, this mutant combination was found in the clinical isolate.TEM-52 (Poyart *et al.*, 42 Antimicrob. Agents Chemother. 108-13 (1998)). This triple mutant was also obtained by directed evolution utilizing the DNA analogs method (Zaccolo (1999)), and additionally was obtained in the present study using *in vivo* shuffling with an *E. coli* hypermutator strain. The unexpected identification of the E104K/M182T/G238S combination of mutations using three different directed evolution techniques, and the additional isolation of this triple mutant in nature (TEM-52) highlights the importance of these residue positions in developing an antibiotic resistance phenotype and suggests that a preferred pathway exists for developing an increased resistance to β-lactam antibiotics.

For instance, Fig. 1 is a graph depicting the frequency of mutations obtained depending on the directed evolution methodology employed as compared to natural isolates, and illustrates that directed evolution can identify the same β-lactamase extended-spectrum antibiotic resistant mutations as are observed in nature (clinical settings). Eight of the eleven mutations obtained with directed evolution have been found in nature (indicated with arrows). The

highest frequency naturally occurring mutations at positions 104, 164, 238 and 240 have all been identified by directed evolution experiments. In addition, three different directed evolution methods identified the E104K/M182T/G238S combination of mutations, which is also present in the clinical isolate TEM-52 as discussed above. Frequencies of clinical mutations were derived from the data provided in the Jacoby website database at http://www.lahey.org/studies/webt.htm (2000). Frequency values were derived by determining the number of times a specific mutation appeared in the extended-spectrum β-lactamase mutant population and dividing by the total number of extended spectrum β-lactamase mutants (56 total; for this number, there were 80 mutants total minus 24 inhibitor resistant mutants). Directed evolution frequencies were calculated by determining the number of occurrences of a specific mutation in the experiments analyzed, divided by the total number of isolates analyzed.

Fig. 2 is a time-line illustrating trends that have been observed in β-lactamase mutation combinations and the order of mutation appearance. As mentioned above, a DNA shuffling experiment in 1994 predicted the mutation combination E104K/M182T/G238S as well as predicting mutations at positions 42 and 92 before their appearance in the clinic (these mutations were observed in 1998 (triple mutant), 1996 (position 42 mutant) and 1999 (position 92 mutant)). The evolution of clinically isolated resistance mutants closely mimics the pathway described by the three directed evolution studies.

In addition to mutational 'hotspot' identification (Fig. 1), another powerful use of directed evolution studies is prediction of new, hitherto unobserved mechanisms that can lead to drug resistance. Such mechanistic predictions permit more effective design of antibiotics and design of inhibitors with a reduced tendency to developing resistance to them. According to the present invention, such mechanistic predictions can be made by isolating and studying the order of individual structural and functional changes produced by a sequential mutation and selection scheme.

Thus, according to the present invention, a number of mutations have been found to characterize the extended substrate spectrum of TEM-type lactamases. An R164S or R164H substitution leads to high resistance to the antibiotic ceftazadime, but low levels of resistance to cefotaxime (Jacoby *et al.*(2000); Giakkoupi *et al.*, 45 J. Antimicrob. Chemother. 101-04 (2000); Knox, 39 Antimicrob. Agents Chemother. 2593-2601 (1995); Matagne *et al.*, 330 Biochem. J. 581-98 (1998)). The G238S mutation confers high resistance to both ceftazadime and cefotaxime, whereas mutation at either position 69 or 244

confers resistance to β-lactamase inhibitors such as tazobactam or clavulanate (Knox (1995)). However, the R164S and G238S combination of mutations are deleterious to extended-spectrum β-lactamase activity (Giakkoupi *et al.* (2000)) and are only encountered in one natural isolate, and only then in the presence of the stabilizing mutation E104K (Jacoby *et al.* (2000); Raquet *et al.*, 23 Proteins 63-72 (1995)).

In a scheme proposed by Giakkoupi *et al.* (2000) and Du Bois *et al.*, 35 J. Antimicrob. Chemother. 7-22 (1995), the most likely sequence of mutational events is first R164S, then E104K, and then G238S. In fact, this trend was detected by the DNA analogs experiment using different levels of cefotaxime selection (Zaccolo *et al.*, (1999)). Low levels of selection produced a R164S mutation, increased levels of selection pressure produced the R164S/E104K combination of mutations, and further mutation of this isolate combined with selection at high cefotaxime concentrations resulted in position 164 reverting to wild type and the isolation of the E104K/M182T/G238S combination of mutations (TEM-52) (Fig. 2).

In addition, Long-McGie *et al.*, 68 Biotechnol. Bioeng. 121-25 (2000) reported on a related phagemid-based *in vivo* directed evolution method using TEM-1 β-lactamase, where they obtained a triple mutant similar to TEM-52. Their mutant contained the same two early mutations as found in TEM-52 (E104K/G238S) along with a third mutation at T265M, allowing for a mutant displaying a similar increased cefotaxime resistance. These results lend further support to the importance of mutational hotspots in the evolution of extended-spectrum antibiotic resistant β-lactamases.

DNA shuffling has further shown that the addition of one signal sequence mutation (A18V) to TEM-52 increased cefotaxime resistance by 16,000-fold (Stemmer (1994)). Three substitutions (A42G, G92S and R241H) to TEM-52 were found to increase resistance 32,000-fold. This ensemble of six mutations has not yet been isolated in nature, but an alarming observation is that individual mutations at five of the six positions have been observed in different clinical isolates. Residue positions 42, 92 and 241 are all located in loop regions of TEM-52, and possibly act to stabilize the reorganized active site shape (Fig. 3a). The stepwise accumulation of mutations observed during the emergence of extended-spectrum antibiotic resistance makes it highly probable that higher order combinations of the predicted mutations will be clinically isolated. Therefore, understanding the development and mechanism of antibiotic resistance conferred by the TEM-52 mutations, as well as determining the

location and role of future mutations predicted to be incorporated in the TEM-52 enzyme architecture, are of utmost importance.

Structural Analysis

Any method known to one of skill in the art for localizing an active site or a functional domain within a protein may be used. For example, the primary amino acid sequence of a protein can be ascertained by DNA and/or protein sequencing methods. The amino acids which are present in the active site are those that play a role in the function of the target protein. Such key amino acids can be identified by observing which amino acids are responsible for phenotypic changes when the target protein is mutated. Similarly, homologies to known "consensus" protein sequences can identify an active site of a target protein. Thus, the primary amino acid sequence of a target protein can be compared to that of known, conserved amino acid sequences which have a known function. Homologous target protein domains can have the same function as that of the "consensus" protein domain.

However, according to the present invention, in addition to information about the primary amino acid sequence, three-dimensional structural information is also needed to optimally design compounds which can interact with and modulate the activity of the target protein. For three-dimensional structural analysis, proteins may be expressed and purified using any technique known to one of skill in the art. For example, see Sambrook, J. *et al.*, Molecular Cloning: a Laboratory Manual, pp. 17.37-17.41, Cold Spring Harbor Laboratory Press (1989); and Rudolph, R. *et al.*, FASEB J. 10:49-56 (1995). Three-dimensional structural analysis can also be performed by any method known to one of skill in the art. For example, such three-dimensional structural analysis can be performed by NMR spectroscopy, electron crystallography or X-ray diffraction crystal analysis. Preferred structural analysis is by X-ray diffraction crystal analysis, with crystals effectively diffracting X-rays for the determination of atomic coordinates to a resolution of better than about 5.0 Angstroms, and more preferably better than about 3.5 Angstroms, and even more preferably better than about 2.5 to about 2.75 Angstroms.

Perhaps the best known way of determining protein structure is X-ray crystallography. See Physical Biochemistry, Van Holde, K. E. (Prentice-Hall,

New Jersey 1971), pp. 221-239, and Physical Chemistry with Applications to the Life Sciences, D. Eisenberg & D. C. Crothers (Benjamin Cummings, Menlo Park 1979). With X-ray crystallography it is possible to elucidate three-dimensional structure with good precision. Protein structures can also be determined by neutron diffraction and nuclear magnetic resonance (NMR). See, e.g., Physical Chemistry, 4th Ed. Moore, W. J. (Prentice-Hall, New Jersey 1972), and NMR of Proteins and Nucleic Acids, K. Wuthrich (Wiley-Interscience, New York 1986). Proteins may also be electrocrystallized for a determination of three-dimensional structure, for instance as described in U.S. Patent 5,597,457, which is herein incorporated by reference in its entirety. A preferred method of crystallization may be found in U.S. Patent No. 6,296,673 which is herein incorporated by reference in its entirety.

The three-dimensional structure of some proteins may also be obtained from public databases. For instance, protein structures of atomic resolution are catalogued by Brookhaven National Laboratory and made available to the public via the Internet, *e.g.*, at http://www.rcsb.org. The Protein Data Bank (PDB) is an archival computer database of three-dimensional structures of biological macromolecules. The database contains atomic coordinates, bibliographic citations, primary sequence and secondary structure information, as well as crystallographic structure factors and 2D-NMR experimental data. Information is available on protein, DNA, RNA, virus and carbohydrate structures. For instance, the atomic coordinates of beta-lactamase TEM52 are available under PDB Accession code 1HTZ, which is herein incorporated by reference.

In summary, the structure of the drug resistant target proteins and variants can be accomplished as follows.

Single crystals of each drug resistant protein are be mounted on an x-ray source such as a rotating anode x-ray generator or a synchrotron. Diffraction images of the protein in question will then be collected utilizing the technique of oscillation photography. The data collected by the above method will then be processed to determine crystal symmetry, unit cell parameters, and other necessary data collection parameters with the programs Denzo, Mosflm or the D*trek suite.

Once the above parameters have been determined and refined, Denzo, Mosflm or the D*trek suite, will be used to measure the integrated intensity of all the reciprocal lattice points contained in the diffraction images collected. This data will then be scaled and reduced by combining redundant measurements using any of the programs, Scalepack, Scala, or the D*trek program dtscaleaverage. In this way a list of reciprocal lattice points and their associated miller indices will be generated along with each points average integrated intensity and associated error. This list of data points will then be processed further, for example, by using the CCP4 suite of crystallography programs. In this case the data will be passed through the program TRUNCATE to convert integrated intensities to structure factor amplitudes, then through the program CAD to ensure the data represents the correct area of reciprocal lattice space.

Phase information will be obtained in the following manner. If a structure homologous to the one being solved exists phase information will be obtained by the method of molecular replacement. In this method the orientation and position of a search model will be optimized in order to achieve the best possible fit between data derived from the search model and experimental data collected above. This procedure will be carried out with any one or more of the following programs: AMORE or MOLREP in the CCP4 suite, or with programs such as EPMR or any other program capable of fitting search model to experimental crystallographic data. Once obtained the phases will be refined by first building a molecular model of the protein structure into the observed data either manually with the program O or XTALVIEW, or automatically with programs such as ARP/WARP or SOLVE/RESOLVE. Phases will be refined further by refining the molecular model against crystallographic, geometric, and chemical data using programs such as REFMAC, CNS, or SHELX.

Drug Design

The present invention provides methods for identifying new drugs, antibiotics, inhibitors and modulators of target antibiotic-resistance proteins that are characterized by their three-dimensional structures. Such drugs, antibiotics, modulators and inhibitors may be competitive, *i.e.*, binding to all or a portion of the active site of the target proteins, or non-competitive, *i.e.*, binding to and inhibiting the target protein whether or not it is bound to another chemical entity.

Drugs and new drugs can act directly or indirectly on a target protein or variant to prevent or reduce drug resistance caused by that target protein.

Antibiotics are drugs that have cytostatic or cytotoxic effects on target organisms. The key to success for an antibiotic is selectivity for the disease target, and lack of toxicity to the patient. The most useful antibiotics against infections are those which attack a microbe-specific target as discussed above.

One structure-based drug design approach is to probe the target protein crystal with molecules composed of a variety of different chemical entities to determine optimal sites for interaction between candidate antibiotics or inhibitors and target proteins. For example, high resolution X-ray diffraction data collected from crystals saturated with solvent allows the determination of where each type of solvent molecule sticks. Small molecules that bind tightly to those sites can then be designed and synthesized and tested for inhibitory activity (J. Travis, Science, 262:1374 (1993)).

This invention also enables the development of compounds that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that binds to or with a target protein. Thus, the time-dependent analysis of structural changes in a target protein during its interaction with other molecules is permitted. The reaction intermediates of a target protein can also be deduced from analysis of the antibiotic or inhibitor in co-complex with a target protein. Such information is useful to design improved analogues of known antibiotics and target protein inhibitors or to design novel classes of antibiotics and inhibitors based on the reaction intermediates observed with a target protein inhibitor co-complex. This provides a novel route for structure-based drug design of antibiotics and inhibitors with both high specificity and stability.

Another structure-based drug design approach made possible by this invention is the ability to perform virtual ligand screening to computationally screen small molecule data bases for chemical entities or compounds that can bind in whole, or in part, to a target protein. In this screening, the quality of fit of such entities or compounds to the binding site may be judged either by shape complementarity or by estimated interaction energy (E. C. Meng et al, J. Comp. Chem., 13:505-524 (1992)).

Because a target protein may crystallize in more than one crystal form, the structure coordinates of a target protein, or portions thereof, are particularly useful to solve the structure of those other crystal forms of target proteins. They may also be used to solve the structure of a mutant target protein, or to solve the

structure of the crystalline form of any other protein with significant amino acid sequence homology to any functional domain of such a target protein.

One method that may be employed for this purpose is molecular replacement. In this method, the unknown crystal structure, whether it is another crystal form of a target protein, or a mutant target protein, or an antibiotic-target protein co-complex, or a crystal of some other protein with significant amino acid sequence homology to any functional domain of a target protein, may be determined by using the coordinates of related known target proteins. This method will provide an accurate structural form for the unknown crystal more quickly and efficiently than attempting to determine such information *ab initio*.

Thus, known three-dimensional structures permit structure-based drug design via screening of known molecules and/or designing of new molecules which bind to a target protein structure, particularly at the active site, via the use of computerized evaluation systems. For example, computer modeling systems are available in which the sequence and structure of a target protein (*i.e.*, the bond angles, dihedral angles, distances between atoms in the active site region, etc.) may be input. Thus, a machine readable medium may be encoded with data representing the coordinates of a target protein or a target protein active site. The computer then generates structural details of the site into which a test compound should bind, thereby enabling the determination of the complementary structural details of the test compound.

The invention therefore provides a computer-assisted method **600** for identifying a new drug effective against at least one drug-resistant variant of a target protein (see Fig. 6). The method generally involves obtaining the three dimensional structure of at least one variant site of interaction with the drug by using computer processing to generate x, y and z coordinates for each atom in the variant site of interaction **602**. The three dimensional structure of the target protein site of interaction with the drug is also obtained by using computer processing to generate an x, y and z coordinates for each atom in the target protein site of interaction **604**. The three dimensional structure of the target protein site of interaction may be available and need not be obtained every time a new variant is isolated. Another step in the computer-assisted method **600** is measuring the differences in distance between the x, y and z coordinates for each atom in the variant and target protein sites of interaction **606**. Such measurements permit one of skill in the art to understand how the drug interaction site may have changed in response to the drug and how to better design a new drug. Hence, the last step in the computer-assisted method **600** is designing a new drug to interact with the site of interaction of at least one of the

variants by using computer processing to fit a library of potential new drugs into the variant and target protein sites of interaction and determining which potential new drug fits into the variant site of interaction better than into the target protein site of interaction **608**.

The invention also provides unique combinations of data fields and data structures useful for practicing the invention (see e.g. Fig. 5). One example of such a data structure **500** stores information on directed evolution of a series of drug resistant variants. This data structure **500** comprises a series of separate variant data fields **502**, each variant data field comprising each x, y and z atomic coordinate of each drug resistant variant. The data structure **500** also includes a target protein data field **504** comprising each x, y and z atomic coordinate of the target protein. The data structure **500** can further include a structural difference data field **506** comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein. The data structure **500** can also have a series of drug resistance data fields **508**, each drug resistance data field comprising each x, y and z atomic coordinate of each drug that is resistant to a drug resistance variant.

The invention also provides an article of manufacture having instructions stored on it that cause a computing system (e.g. **702**) to identify a new drug effective against at least one drug-resistant variant of a target protein, the instructions comprising: instructions for comparing a series of separate variant data fields with a target protein data field to generate a structural difference data field comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein, wherein each variant data field within the series of separate variant data fields comprises each x, y and z atomic coordinate of each drug resistant variant, and wherein a target protein data field comprises each x, y and z atomic coordinate of the target protein; instructions for generating a new drug data field from the structural difference data field and a small molecule data field, wherein the small molecule data field comprises x, y and z atomic coordinates for a series of small molecules and wherein the new drug data field comprises small molecules that can optimally fit within the site of interaction of the drug-resistant variant. New drugs identified using the article of manufacture are also provided by the invention.

Examples of small molecule inhibitors that were designed based on protein three-dimensional structure information are available for HIV protease (L. Hong *et al.*, Protein Sci. 9:1898-1904 (2000); J.V.N.V. Prasad *et al.*, Bioorg. Med. Chem. 7:2775-2800 (1999); A. Wlodawer *et al.*, Annu. Rep. Biophys.

Biomol. Struct. 27:249-84 (1998)), renin (J. Rahuel *et al.*, Chem. Biol. 7:493-504 (2000)), Factor Xa (S. Maiganan *et al.*, Curr. Topics Med. Chem. 1:161-74 (2001)), thrombin (N.Y. Chirgadze *et al.*, Protein Sci. 9:29-36 (2000), B.A. Katz *et al.*, J. Mol. Biol. 307:1451-86 (2001)), urokinase (V.L. Nienaber *et al.*, Structure 8:553-63 (2000)), rhinovirus 3C protease (S.H. Reich *et al.*, J. Med. Chem. 43:1670-1683 (2000)), cathepsin K (R.W. Marquis *et al.*, J. Med. Chem. 44:725-36 (2001)), Abl kinase (T. Schindler *et al.*, Science 289:1938-42 (2000)), CDKs (T.G. Davies *et al.*, Structure 9:389-97 (2001), P. Furet *et al.*, J. Comput. Aided Mol. Des. 14:403-9 (2000), N.S. Gray *et al.*, Science 281:533-8 (1998), S.-H. Kim Pure Appl. Chem. 70:555-65 (1998)), EGFR kinase (S. Ghosh *et al.*, Curr. Cancer Drug Targets 1:129-40 (2001)), Lck kinase (X. Zhu *et al.*, Structure 7:651-61 (1999)), Src kinase (T.K. Sawyer *et al.*, Expert Opin. Invest. Drugs 10:1327-44 (2001), W.C. Shakespeare *et al.*, Proc. Natl. Acad. Sci. USA 15:9373-8 (2000), W. Xu *et al.*, Mol. Cell 3:629-36 (1999)), as well as aldose reductase (Y. Iwata *et al.*, J. Med. Chem. 44:1718-28 (2001)), phospholipase-A2 (E.D. Mihelich *et al.*, Biochim. Biophys. Acta 1441:223-8 (1999)), STAT (J. Hurkson *et al.*, Oncogene 19:6594-9 (2000)), Bcl-2 (Z. Huang Oncogene 19:6627-31 (2000)), neuraminidase (Y.S. Babu *et al.*, J. Med. Chem. 43:3482-6 (2000)), and glyceraldehyde-3-phosphate dehydrogenase (J.C. Bressi *et al.*, J. Med. Chem. 44:2080-93 (2001)).

More particularly, the design of drugs and compounds that bind to or inhibit a target protein according to this invention generally involves consideration of two factors. First, the drug or compound must be capable of physically and structurally associating with a target protein. Non-covalent molecular interactions important in the association of a target protein with antibiotics and inhibitors include hydrogen bonding, van der Waals, hydrophobic and/or covalent interactions. Second, the drug or compound must be able to assume a conformation that allows it to associate with a target protein. Although certain portions of the drug or compound will not directly participate in this association with a target protein, those portions may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on potency. Such conformational requirements include the overall three-dimensional structure and orientation of the chemical entity or compound in relation to all or a portion of the binding site, e.g., active site or accessory binding site of a target protein, or the spacing between functional groups of a compound comprising several chemical entities that directly interact with a target protein.

The potential inhibitory or binding effect of a drug or compound on a target protein may be analyzed prior to its actual synthesis and testing by the use of computer modeling techniques. If the theoretical structure of the given drug or compound suggests insufficient interaction and association between it and a target protein, synthesis and testing of the compound is obviated. However, if computer modeling indicates a strong interaction, the molecule may then be synthesized and tested for its ability to bind to a target protein and inhibit using a suitable assay. In this manner, synthesis of inoperative compounds may be avoided.

Compounds and drugs capable of binding to a target protein of the present invention may be computationally evaluated and designed by means of a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with the individual binding pockets or other areas of the target protein, in the mutant and wild-type forms of the protein. Once a three-dimensional structure of a biological macromolecule (target) has been determined by X-ray, NMR, or modeling, one can then perform structure-based drug design to computationally fit millions of small molecule compounds into complimentary regions of the target. These regions of the target structure can be active sites, allosteric sites, regulatory sites, global suppressor sites, partner binding sites, *etc.*

One skilled in the art may use one of several structure-based drug design methods to computationally screen chemical entities or fragments for their ability to associate with a target protein and more particularly with the individual binding pockets of the target protein active site or accessory binding site. This process may begin by visual inspection of, for example, the active site on the computer screen based on available target protein coordinates. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within a binding pocket of a target protein, in the mutant and wild-type forms of the protein. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics calculations with standard molecular mechanics force fields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the structure-based drug design process of selecting fragments or chemical entities. These include:

GRID [P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28:849-857 (1985)]. GRID is available from Oxford University, Oxford, UK.

MCSS [A. Miranker and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method", Proteins: Structure, Function and Genetics, 11:29-34 (1991)]. MCSS is available from Molecular Simulations, Burlington, Mass.

AUTODOCK [D. S. Goodsell and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8:195-202 (1990)]. AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.

DOCK [I. D. Kuntz et al, "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288 (1982)]. DOCK is available from University of California, San Francisco, Calif.

Once suitable chemical entities or fragments have been selected, they can be assembled into a single antibiotic or inhibitor compound. Visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of the target proteins is followed by manual model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

CAVEAT [P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc. 78, pp. 182-196 (1989)]. CAVEAT is available from the University of California, Berkeley, Calif.

3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, Calif.). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35:2145-2154 (1992).

HOOK (available from Molecular Simulations, Burlington, Mass.).

Instead of proceeding to build a target protein inhibitor or antibiotic in a step-wise fashion, building up from one fragment or chemical entity at a time as described above, inhibitory or antibiotic compounds may be designed as a whole or "de novo" using either an empty active site or optionally including some portion(s) of a known inhibitor(s). These methods include:

LUDI [H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid.

Molec. Design, 6:61-78 (1992)]. LUDI is available from Biosym Technologies, San Diego, Calif.

LEGEND [Y. Nishibata and A. Itai, Tetrahedron, 47:8985 (1991)].

LEGEND is available from Molecular Simulations, Burlington, Mass.

LEAPFROG (available from Tripos Associates, St. Louis, Mo.).

Other molecular modeling techniques may also be employed in accordance with this invention. See, e.g., N. C. Cohen et al, "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33:883-894 (1990). See also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2:202-210 (1992). For example, where the structures of test compounds are known, a model of the test compound may be superimposed over the model of the structure of the invention. Numerous methods and techniques are known in the art for performing this step, any of which may be used. See, *e.g.*, P. S. Farmer, Drug Design, Ariens, E. J., ed., Vol. 10, pp 119-143 (Academic Press, New York, 1980); U.S. Pat. No. 5,331,573; U.S. Pat. No. 5,500,807; C. Verlinde, Curr. Biol., 2:577-587 (1994); and I. D. Kuntz, Science, 257:1078-1082 (1992). The model building techniques and computer evaluation systems described herein are not a limitation on the present invention.

Thus, using these computer evaluation systems, a large number of compounds may be quickly and easily examined, and expensive and lengthy testing avoided. Moreover, the need for actual synthesis of many compounds is effectively eliminated.

Another embodiment of the present invention involves the use of high-throughput ligand screening methods to identify small molecules capable of binding or interacting with a target. Once a mutant or wild type target protein is isolated (purified), or expressed in a cell line that is coupled to any readout, hundred to thousands (to millions for ultra high-throughput) of small molecules can be screened against the target molecule for detecting a change in a biophysical property of the target or the readout molecule. Readouts include radioactivity, fluorescence, colorimetric, *etc.*

Once identified by the modeling or screening techniques, the antibiotic or inhibitor or antibiotic-resistance inhibitor may be tested for bioactivity using standard techniques. For example, test antibiotics and inhibitors can be tested in binding assays using conventional formats to screen for activity or binding. One particularly suitable assay format includes the enzyme-linked immunosorbent

assay (ELISA). Other assay formats may be used; these assay formats are not a limitation on the present invention.

In another aspect of the present invention, target protein structures permit the design and identification of synthetic compounds and/or other molecules which are characterized by the conformation of the target proteins. Using known computer systems, the coordinates of the structures of the invention may be provided in machine readable form, the test antibiotics and inhibitors designed and/or screened, and their conformations superimposed on the structures of the target proteins of the invention. Subsequently, suitable candidates identified as above may be screened for the desired inhibitory activity, stability, and the like.

An additional embodiment of the present invention involves the use of Structure Activity Relationships to develop, design, or refine candidate small molecule inhibitors or antibiotics. Using a variety of biophysical techniques, such as screening, CD, NMR, X-ray, modelling, Mass spec, etc..., structure activity relationships can be determined for any target molecule against a variety of small molecule inhibitors (Tollenaere, J.P. Pharm. World Sci. 18(2): 56-62 (1996)). In drug design, structure-activity relationships (SAR) use the relationship between chemical structure and pharmacological activity for a series of compounds to derive and design improved drug species. Another algorithm used is Quantitative Structure-Activity Relationships (QSAR), which are mathematical relationships linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. Methods which can be used in QSAR include various regression and pattern recognition techniques. In general, structure-property correlations (SPC) refer to all statistical mathematical methods that are used to correlate any structural property to any other property (intrinsic, chemical or biological), using statistical regression and pattern recognition techniques. Finally, three-dimensional quantitative structure-activity relationships (3D-QSAR) are the analysis of the quantitative relationship between the biological activity of a set of compounds and their spatial properties using statistical methods. (all definitions from http://www.chem.qmw.ac.uk/iupac/medchem/ix.html#s5). At http://www.pharmacopeia, SAR is defined as an analysis which defines the relationship between the structure of a molecule and its ability to affect a biological system (be that a cell, protein, enzyme, or organism).

The invention also provides a system **700** that includes a computer system **702**; memory **704** accessible to the computer system **702**, a series of data structures **706** stored in the memory and one or more software components operable on the computer system. A computer system **702** is any type of device having at least a processor and accessible memory **704** and associated networks and peripherals, such as local area networks (LANS) and printers. See e.g., Fig. 7. Examples of computers include personal computers, laptops, supercomputers, and any other kind of computer. The computer system **702** is capable of operating according to instructions, software components, and data structures.

In particular, the invention provides a system **700** comprising a computer system **702**; memory **704** accessible to the computer system **702**; a first data structure **706** stored in the memory **704** comprising three dimensional structural data of a target protein; a second data structure **706** stored in the memory **704** comprising a first population of three dimensional structural data for a series of variants of the target protein, wherein the variants were generated by one or more rounds of directed evolution of the target protein in the presence of a drug; a third data structure **706** stored in the memory **704** comprising a second population of three dimensional structural data for a series of second generation variants of a first variant, wherein the second generation of variants is generated by one or more rounds of directed evolution of the first variant in the presence of a drug; a fourth data structure **706** comprising three dimensional structural data for a population of potential new drugs comprising antibiotics, enzyme inhibitors, protein inhibitors, antibiotic resistance inhibitors, herbicide resistance inhibitors, insect repellants, insecticide resistance inhibitors, viral drug resistance inhibitors, chemotherapeutic resistance inhibitors, modulators, antagonists, agonists, effectors, ligands, antibodies, antibody fragments, peptides or small molecules; a first software component **706** operable on the computer system **702** to compare the first data structure with the second or third data structure and to identify three dimensional structural variants that represent a drug resistance evolution profile of said target protein; a second software component **706** operable on the computer system **702** to compare the fourth data structure **706** with the drug resistance evolution profile of said target protein and thereby to design a new drug effective against at least one drug resistance variant.

Once identified and screened for biological activity, these new drugs, antibiotics, inhibitors and modulators may be used therapeutically or prophylactically to block target protein activity, and thus, cellular or viral viability.

EXAMPLES

The following examples are provided to describe and illustrate the present invention. As such, they should not be construed to limit the scope of the invention. Those in the art will well appreciate that many other embodiments also fall within the scope of the invention, as it is described hereinabove and in the claims.

Example 1: Beta-Lactamase (TEM52) Penicillins

According to the present invention, antibiotic resistance involves the accumulation of mutations beneficial to the protein while maintaining residue interactions and core packings that are critical for preserving function. The constraints of maintaining stability, while increasing activity, drastically reduces the number of possible mutational combination pathways. To test this theory, the evolution of TEM-1 β-lactamase was directed using a hypermutator *E. coli*-based directed evolution technique with cefotaxime selection. The selected mutants were compared to two previous directed evolution studies and a database of clinical isolates. In all cases, evolution resulted in the generation of the E104K/M182T/G238S combination of mutations with 500-fold increased resistance, which is equivalent to clinical isolate TEM-52. The structure of TEM-52 was determined to 2.4 Å. The G238S mutation widens access to the β-lactamase active site by 2.8 Å whereas the E104K mutation stabilizes the reorganized topology of β-lactamase. The M182T mutation is located 17 Å from the active site and appears to be a global suppressor mutation that acts to stabilize the new enzyme structure.

Methods

Mutator strain directed evolution. Each cycle of *in vivo* DNA shuffling of β-lactamase consisted of four steps: (i) infection, mutagenesis, and packaging of a TEM-1 β-lactamase-encoding phagemid, pBAD18 (ATCC), in the hypermutator strain MG163 (generated by phage P1 transduction of mutD5 (Echols *et al.*, 80 Proc. Nat'l Acad. Sci. 2189-92 (1983); Low *et al.*, 260 J. Mol. Biol. 359-68 (1996)) from donor strain DM2516 (a gift of R.A. Kelln) into the

F^+ $recA^+$ strain NM522 via the linked transposon zaf-13::Tn10); (ii) superinfection of the shuffler strain MG109 (generated by transduction of strain NM522 with a P1 lysate grown on a mutS::Tn5 donor (a gift of D. Thayler)) with the mutagenized phagemid pool for recombination; (iii) selection on increasing levels of cefotaxime (Sigma); and (iv) pooling resistant colonies and rescuing phagemid for the next round of shuffling. Substitution of wild-type strain NM522 in the *in vivo* DNA shuffling protocol was used as a non-mutator control.

Protein production, crystallization and structure determination. The TEM-52 gene was PCR amplified (primers: 5'-ATGAGTATTCAACATATGTTCCTGGTCGCC-3' (SEQ ID NO:1) and 5'-TTACCAATGCTTGGATCCAATCAGTGAGGC-3' (SEQ ID NO:2)), cloned into pET30a with *Nde*I and *Bam*HI, and transformed into BL21(DE3) pLysS. The DNA sequence of the TEM-52 open reading frame is as follows (SEQ ID NO:3):

ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGG
CATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAG
TAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATC
GAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC
CGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATG
TGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCG
GTCGCCGCATACACTATTCTCAGAATGACTTGGTTAAGTACTCAC
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAA
TTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAAC
TTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTT
TTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAA
CCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAC
GACGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGG
CGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTC
CGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCAGTGAGCGTG
GGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCT
CCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGG
ATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA
AGCATTGGTAA

The TEM-52 protein was purified according to the method described in Jelsch *et al.*, 16 Proteins 364-83 (1993). The sequence of the TEM-52 protein is as follows (SEQ ID NO:4).

MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYIEL
DLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIH
YSQNDLVKYSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTI
GGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTTPAAMA
TTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIAD
KSGASERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIG
ASLIKHW

The TEM-52 protein was crystallized by hanging drop method (14 mg/ ml) over 16% PEG 8K,125 mM MgOAc, 25 mM Na Citrate, pH 6.5. Crystals appeared after 3 days; space group $P4_32_12$ (a = b = 88.4 Å, c = 500.4 Å), six molecules per asymmetric unit.

Data were collected to 2.4 Å on three frozen crystals at the Advanced Light Source beamline 5.0.2, and processed with DENZO and SCALEPACK (Otwinowski *et al.*, 276 Methods Enzymol. 307-26 (1997)). Because of the long c-axis, several attempts were made on three crystals in different orientations to collect a missing cone of data and completeness up to 90% was obtained. In addition, attempts were made to process the data in a hexagonal or rhombohedral lattice since a hexamer was observed in the tetragonal space group. However, all attempts at processing in the higher symmetry were unsuccessful. Molecular replacement with 1TEM as the model (Jelsch *et al.*, 16 Proteins 364-83 (1993)) using the program AMORE (Navaza 50 Acta Crystallogr. A157-63 (1994)) located five of the six molecules. An $F_o - F_c$ map revealed density for a sixth molecule. All six molecules were rebuilt using O(Jones *et al.*, 47 Acta Crystallogr. A 110-19 (1991)) and refined using CNS (Brunger *et al.*, 54 Acta. Crystallogr. D 905-21 (1998)) until convergence at 2.4 Å with a R_{cryst} = 21.7 and R_{free} = 26.1. Five of the six molecules adhere to non-crystallographic symmetry, with the sixth molecule being very similar in structure, but not able to refine with NCS due to packing differences. The coordinates have been deposited in the Protein Data Bank (accession code 1HTZ); which are herein incorporated by reference in their entirety.

Table 2 Data Collection and Refinement Statistics

Data Collection	
Resolution (Å)	20–2.40
Molecules per asymmetric unit	6
Number of reflections	736,906
Number of unique reflections	73,211
Redundancy [1]	10 (6)
I/σ [1]	22.4 (7.1)
R_{symm} (%) [1,2]	11.5 (43.2)
Completeness (%) [1]	91 (80)
Refinement (for all six molecules)	
Resolution (Å)	20–2.40
Reflections	66,385
Wilson B-factor ($Å^2$)	43.5
R_{cryst} (%) [3] / R_{free} (%) [4]	21.7 / 26.1
Protein Atoms	10,140
Number of waters	587
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.3
Average r.m.s.deviation [5]	
Residues 100–110	0.49
Residues 237–244	0.35

[1] Numbers in parenthesis denote values for highest resolution bin (2.46–2.40 Å).

[2] FIX EQUATION$R_{symm} = \Sigma| I_{hi} - <I_h>| / \Sigma I_{hi}$ for all h, where I_h is the intensity of reflection h.

[3] FIX EQUATION$R_{cryst} = \Sigma|| F_o |- |F|| / \Sigma|| F_o |$.

[4] R_{free} was calculated with 10% of the reflections not included in refinement. All reflections were used in the final round of refinement.

[5] Noncrystallographic symmetry (NCS) was applied to five of the six molecules. Restraints were left off the sixth molecule because of the differences caused by crystal packing. In the final round of building and refinement NCS was removed for residues 100–110 and 237–244.

TEM-52 Structure

To understand the role of the mutations E104K/M182T/G238S in natural and directed evolution, the crystal structure of TEM-52 was determined to 2.4 Å (Figs. 3, 4 and 5). Fig. 3 provides alternative views of the TEM-52 three-dimensional structure. Fig. 3a illustrates that both nature and directed evolution experiments identified the E104K/M182T/G238S combination of mutations (green/red/green) and DNA shuffling predicted the loop mutations at positions 42, 92 and 241 (red). Catalytic residue positions are black and the B3 β-strand is red. Fig. 3b provides an overlay stereoview of wild type TEM-1 (cyan) and

TEM-52 (blue) illustrating movement of loops 238–243, 267–271 and 40–43. The G238S mutation (green) causes the Glu 240 side chain conformation to change (compare blue and purple side chains), which widens the active site by 2.8 Å, potentially allowing binding of bulky cephalosporins (catalytic Ser 70 side chain is shown in yellow). Fig. 3c provides a stereoview of the electron density for the TEM-52 active site.

The active site is located at the interface of the two β-lactamase domains. One domain consists of a five-stranded πΣωΠsheet packed against three α-helices, while the second domain consists of eight α-helices. The overall TEM-52 structure is very similar to wild type β-lactamase (TEM-1) with a root mean square (r.m.s.) deviation of 0.66 Å comparing the CALPHA traces. The mutations generated in early rounds of selection (at positions 104 and 238) are involved in reorganizing the shape of the active site (Fig. 3b) while a mutation generated in later stages of the selection (position 182) may compensate for the destabilizing effects introduced by the initial mutations.

An individual G238S mutation results in an eight-fold increase in minimum inhibitory concentration (MIC) (*see* Stemmer, 370 Nature 389-91 (1994)), but the single mutant was found to be thermally destabilized (*see* Raquet *et al.*, 23 Proteins 63-72 (1995)). In TEM-52, Ser 238 forms two new hydrogen bonds from the hydroxyl group: one to the Ser 243 backbone amide (2.9 Å, average over six molecules) and one to the Ser 243 hydroxyl group (3.2 Å) (Fig. 3 b,c). Loop 238–243 that contains this mutation is altered by as much as 2.8 Å (measured by the CALPHA positions of Glu 240 for TEM-52 versus TEM-1), widening the opening to the active site and potentially facilitating binding of bulky cephalosporin antibiotics. Movement of loop 238–243 results in the reorganization of adjacent loops 267–271 and 40–43. In addition, loop 238–243 repositioning causes the backbone carbonyl of Glu 240, which was hydrogen bonded with the Ser 243 backbone amide in TEM-1, to rotate 155° in TEM-52, forming a hydrogen bond with the Gly 242 backbone amide group. In TEM-1, the Glu 240 carboxylate group was thought to hinder binding of some cephalosporins and β-lactamase inhibitors by steric repulsion (*see* Medeiros, 24 Clin. Infect. Dis. (Suppl. 1), S19-45 (1997)). However, in the TEM-52 structure this side chain is repositioned out of the active site and oriented over the B3 πΣωΠ-strand, reducing the possibility of any interference with substrate or inhibitor binding.

The second TEM-52 mutation, E104K, when observed as a single mutation in the wild type enzyme background, is found to increase the MIC for cefotaxime four-fold (*see* Stemmer, 370 Nature 389-91 (1994); Raquet *et al.*, 23

Proteins 63-72 (1995)). This mutation acts synergistically with the G238S mutation, with the double mutant displaying a 267-fold increase in MIC *versus* TEM-1 (*see* Stemmer, 370 Nature 389-91 (1994)). In the crystal structure of TEM-52, for the six molecules that are located in the asymmetric unit, all six Lys 104 residues have well-defined electron density (Fig. 3c). This is particularly striking since Lys 104 is located near the surface, and surface-exposed lysine residues are typically disordered. In TEM-52, Lys 104 from the E104K mutation is oriented into the active site toward the reorganized Glu 240 side chain. In the wild type structure, the Glu 104 and Glu 240 carboxyl groups were separated by 6.8 Å (Vanhove *et al.*, 54 Cell. Mol. Life Sci. 372-77 (1998)). However, in TEM-52, the E104K and Glu 240 side chains are only 4.8 Å apart (Fig. 3c).

M182T is a global suppressor mutation

Perhaps the most interesting mutation found in TEM-52 is M182T, which has the ability to reverse the destabilizing effects of antibiotic- or inhibitor-resistant mutations (*see* Huang *et al.*, 94 Proc. Nat'l. Acad. Sci. 8801-06 (1997); Farzaneh *et al.*, 40 Antimicrob. Agents Chemother. 2434-36 (1996)). M182T is the only mutation found in both antibiotic-resistant and inhibitor-resistant β-lactamases, suggesting a potential role as a global suppressor. The M182T mutation in combination with the E104K/G238S mutations increases the cefotaxime MIC to 500 μg/ml.

Fig. 4 depicts the structure of the region around the global suppressor mutation M182T found in TEM-52. Fig. 4a provides the structure of the β-lactamase M182T mutant (light purple), and identifies the location of this mutation as being in a hinge region (arrow) between two domains and 17 Å from the active site. Thr 182 (light purple) in one domain (blue) forms new interactions with residues 63 (green) and 64 (purple) in the second domain (yellow). Fig. 4b provides an alternative view of TEM-52, illustrating the interactions relevant to the Thr 182 (light purple) mutation. M182T compensates for the destabilizing mutations found at position 69 (inhibitor-resistant mutants) and positions 164 or 238 (antibiotic-resistant mutations) (yellow), potentially contributing to stabilization of the reorganized shape of the active site (catalytic residues shown in red). Small molecules designed to destabilize the region around position 182 may lead to a novel class of antibiotics or inhibitors.

Amino acids leading from the region around position 182 to the catalytic triad do not generally tolerate substitutions, and are thought to be necessary for

proper core packing and catalytic residue orientation (Huang *et al.*, 258 J. Mol. Biol. 688-703 (1996)). Phe 66 (a highly conserved residue in Class A β-lactamases) and Pro 67 are involved in hydrophobic core packing interactions; Thr 71 and Lys 73 are important for proper positioning of the catalytic residues Ser 70 and Asn 132, while Asp 176, Asp 179, Thr 180 and Thr 181 interact with the active site CAPITAL OMEGA-loop that contains the catalytic residue Glu 166. The hydroxyl group of Thr182 forms two new hydrogen bonds with the backbone carbonyls of residues Glu 63 (2.9 Å) and Glu 64 (3.3 Å), increasing the interaction between the two domains and possibly reducing the conformational flexibility of the hinge region between the two domains. The global suppressor M182T has been previously proposed to reverse the destabilizing effects of antibiotic-resistant mutations at positions 238 (Poyart *et al.*, 42 Antimicrob. Agents Chemother. 108-13 (1998)) and possibly position 164 (Yang, 42 Antimicrob. Agents Chemother. 1671-76 (1998)). M182T may also help to stabilize the inhibitor-resistant mutation at position 69 (Medeiros, 24 Clin. Infect. Dis. (Suppl. 1) S19-45 (1997); Huang *et al.*, 94 Proc. Nat'l. Acad. Sci. USA 8801-06 (1997)).

Global suppressor mutations have been described for phage P22 coat protein (Aramli *et al.*, 274 J. Biol. Chem. 22217-24 (1999), several forms of p53 (Nikolova *et al.*, 19 EMBO J. 370-78 (2000)), β-lactamase (Huang *et al.*, 94 Proc. Nat'l. Acad. Sci. USA 8801-06 (1997)), and have been presumed for HIV protease (*id.*). By a mechanism similar to that proposed to restore function to tumorigenic p53 variants (Nikolova *et al.*, 19 EMBO J. 370-78 (2000)), a small molecule targeted to a global suppressor site may act to either stabilize or destabilize the native fold. Isomerization of the Glu166–Pro167 peptide bond to a *cis* conformation in β-lactamase is thought to occur in the last phase of the five phases in the proposed protein folding pathway (Vanhove *et al.*, 54 Cell. Mol. Life Sci. 372-77 (1998)). This isomerization step is most likely the rate-limiting step (time constant of 300 s) in β-lactamase folding, resulting in collapse of the Ω-loop residues onto the main body of the enzyme (Vanhove *et al.*, 54 Cell. Mol. Life Sci. 372-77 (1998)). Without isomerization of the Pro 167 bond, the Ω-loop is prevented from laying down on top of the hinge region, making the M182T global suppressor region more exposed. Until the Ω-loop assumes the correct position, the catalytic residues are improperly positioned, making β-lactamase inactive. A small molecule targeted to the global suppressor region could possibly bind to the inactive β-lactamase intermediate, preventing Ω-loop collapse and the proper assembly of the catalytic residue geometry. As described above, mutations in this region of β-lactamase are not generally

tolerated, and thus, development of resistance by mutations elicited to a bound molecule at this location would be less likely to occur.

Structural consequences of further mutations to TEM-52

Several combinations of mutations have not been observed in the clinic, but are likely to appear if increased levels of cefotaxime are continually used. For example, mutation E240K was identified by the DNA analogs study (Zaccolo *et al.*, 285 J. Mol. Biol. 775-83 (1999)) and has been observed in thirteen naturally occurring isolates (Fig. 1). A lysine at position 240 of TEM-52 would most likely allow this side chain to interact directly with antibiotic substrates (Matagne *et al.*, 330 Biochem. J. 581-98 (1998)). Additionally, the G238S mutation in TEM-52 causes a minor reorganization of loops 267–271 and 40–43. Consistent with this movement, the DNA shuffling experiment was observed to evolve an A42G mutation. A different substitution at this position, A42V, has been observed (Jacoby *et al.*, http://www.lahey.org/studies/webt.htm (2000)) and may fill the space produced by adjacent loop 267–271 movement, potentially providing further stabilization of the enzyme (Fig. 3b). Finally, the DNA shuffling experiment identified an A18V β-lactamase mutation, located in the signal sequence region, providing for an alternative evolutionary mechanism to increase antibiotic resistance without the requirement of any direct changes in catalytic active site structure.

Drug design based on the TEM-52 β-lactamase mutant

Lead small molecule candidates were selected based on docking with the TEM-52 (PDB code 1HTZ) or TEM-1 (PDB code 1TEM) active sites, using the virtual ligand screening (VLS) technology (ICM Main and ICM-Dock) available from MolSoft, LLC. A library of over 200,000 small molecule compounds were computationally screened using ICM-Dock, and a number of lead inhibitors were ligand screened and identified that interact with either TEM52 or TEM1 or both.

EXAMPLE 2: Drug Design Procedures

The present invention provides methods for identifying new drugs such as antibiotics, inhibitors and modulators of target antibiotic-resistance proteins that are characterized by their three-dimensional structures. Such antibiotics and inhibitors may be competitive, *i.e.*, binding to all or a portion of the active site of the target proteins, or non-competitive, *i.e.*, binding to and inhibiting the target protein whether or not it is bound to another chemical entity.

One structure-based drug design approach is to probe the target protein crystal with molecules composed of a variety of different chemical entities to determine optimal sites for interaction between candidate antibiotics or inhibitors and target proteins. For example, high resolution X-ray diffraction data collected from crystals saturated with solvent allows the determination of where each type of solvent molecule sticks. Small molecules that bind tightly to those sites can then be designed and synthesized and tested for inhibitory activity (J. Travis, Science, 262:1374 (1993)).

This invention also enables the development of compounds that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that binds to or with a target protein. Thus, the time-dependent analysis of structural changes in a target protein during its interaction with other molecules is permitted. The reaction intermediates of a target protein can also be deduced from analysis of the antibiotic or inhibitor in co-complex with a target protein. Such information is useful to design improved analogues of known antibiotics and target protein inhibitors or to design novel classes of antibiotics and inhibitors based on the reaction intermediates observed with a target protein inhibitor co-complex. This provides a novel route for structure-based drug design of antibiotics and inhibitors with both high specificity and stability.

Another structure-based drug design approach made possible by this invention is the ability to perform virtual ligand screening to computationally screen small molecule data bases for chemical entities or compounds that can bind in whole, or in part, to a target protein. In this screening, the quality of fit of such entities or compounds to the Once a three-dimensional structure of a biological macromolecule (target) has been determined by X-ray, NMR, or modeling, one can then perform structure-based drug design to computationally fit millions of small molecule compounds into complimentary regions of the target. These regions of the target structure can be active sites, allosteric sites, regulatory sites, global suppressor sites, partner binding sites, *etc.*

One skilled in the art may use one of several structure-based drug design methods to computationally screen chemical entities or fragments for their ability to associate with a target protein and more particularly with the individual

binding pockets of the target protein active site or accessory binding site. This process may begin by visual inspection of, for example, the active site on the computer screen based on available target protein coordinates. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within a binding pocket of a target protein, in the mutant and wild-type forms of the protein. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics calculations with standard molecular mechanics force fields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the structure-based drug design process of selecting fragments or chemical entities. These include:

GRID [P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28:849-857 (1985)]. GRID is available from Oxford University, Oxford, UK.

MCSS [A. Miranker and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method", Proteins: Structure, Function and Genetics, 11:29-34 (1991)]. MCSS is available from Molecular Simulations, Burlington, Mass.

AUTODOCK [D. S. Goodsell and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8:195-202 (1990)]. AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.

DOCK [I. D. Kuntz et al, "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288 (1982)]. DOCK is available from University of California, San Francisco, Calif.

Once suitable chemical entities or fragments have been selected, they can be assembled into a single antibiotic or inhibitor compound. Visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of the target proteins is followed by manual model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

CAVEAT [P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc. 78, pp. 182-196 (1989)]. CAVEAT is available from the University of California, Berkeley, Calif.

3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, Calif.). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35:2145-2154 (1992).

HOOK (available from Molecular Simulations, Burlington, Mass.).

Instead of proceeding to build a target protein inhibitor or antibiotic in a step-wise fashion, building up from one fragment or chemical entity at a time as described above, inhibitory or antibiotic compounds may be designed as a whole or "de novo" using either an empty active site or optionally including some portion(s) of a known inhibitor(s). These methods include:

LUDI [H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid. Molec. Design, 6:61-78 (1992)]. LUDI is available from Biosym Technologies, San Diego, Calif.

LEGEND [Y. Nishibata and A. Itai, Tetrahedron, 47:8985 (1991)].

LEGEND is available from Molecular Simulations, Burlington, Mass.

LEAPFROG (available from Tripos Associates, St. Louis, Mo.).

Other molecular modeling techniques may also be employed in accordance with this invention. See, e.g., N. C. Cohen et al, "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33:883-894 (1990). See also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2:202-210 (1992). For example, where the structures of test compounds are known, a model of the test compound may be superimposed over the model of

the structure of the invention. Numerous methods and techniques are known in the art for performing this step, any of which may be used. See, *e.g.*, P. S. Farmer, Drug Design, Ariens, E. J., ed., Vol. 10, pp 119-143 (Academic Press, New York, 1980); U.S. Pat. No. 5,331,573; U.S. Pat. No. 5,500,807; C. Verlinde, Curr. Biol., 2:577-587 (1994); and I. D. Kuntz, Science, 257:1078-1082 (1992). The model building techniques and computer evaluation systems described herein are not a limitation on the present invention.

Thus, using these computer evaluation systems, a large number of compounds may be quickly and easily examined, and expensive and lengthy testing avoided. Moreover, the need for actual synthesis of many compounds is effectively eliminated.

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

WHAT IS CLAIMED:

1. A method for identifying a new drug effective against at least one drug-resistant variant of a target protein, comprising:

(a) selecting at least one drug-resistant variant of a target protein in the presence of a drug;

(b) obtaining the three dimensional structure of at least one variant site of interaction with the drug;

(c) obtaining the three dimensional structure of the target protein site of interaction with the drug;

(d) comparing the three dimensional structures of the target and variant sites of interaction to identify structural similarities and differences; and

(e) designing a new drug to interact with at least one of the variants by using the structural similarities or differences.

2. The method of claim 1, wherein the three dimensional structures of the target and variant sites of interaction are determined using x-ray crystallography, NMR, or molecular modeling.

3. The method of claim 1 that further comprises contacting the new drug with the target protein or with at least one drug-resistant variant thereof to compare the interaction of the new drug with the target protein and drug-resistant variant thereof.

4. The method of claim 3, wherein the interaction of the new drug with the target protein or with at least one drug-resistant variant thereof is compared by assaying the function of the target protein or the drug-resistant variant thereof in the presence of the drug.

5. The method of claim 3, wherein the interaction of the new drug with the target protein or with at least one drug-resistant variant thereof is compared by co-crystallizing the new drug and the target protein or the drug-resistant variant thereof.

6. The method of claim 1, wherein at least one drug-resistant variant of the target protein is isolated from a patient who is or was undergoing treatment with the drug.

7. The method of claim 1, wherein at least one drug-resistant variant of the target protein is selected by at least one cycle of directed evolution.

8. The method of claim 7, wherein the directed evolution is accomplished by using random mutagenesis, DNA analogs or DNA shuffling.

9. The method of claim 8, wherein the directed evolution is performed in a hypermutator organism.

10. The method of claim 1, wherein the new drug is designed to interact with both the target protein and at least one drug-resistant variant of the target protein.

11. The method of claim 1, wherein the new drug is designed to interact with regions of at least one drug-resistant variant that are structurally different from analogous regions in the target protein.

12. The method of claim 1, wherein the new drug is designed to inhibit the folding of one or more regions of the drug-resistant variant(s) or the target protein.

13. The method of claim 1, wherein the drug is an antibiotic, an enzyme inhibitor, a protein inhibitor, an antibiotic resistance inhibitor, a herbicide resistance inhibitor, an insect repellant, an insecticide resistance inhibitor, a viral drug resistance inhibitor, a chemotherapeutic resistance inhibitor, a modulator, an antagonist, an agonist, an effector, a ligand, an antibody, an antibody fragment, a peptide or a small molecule.

14. The method of claim 13, wherein said modulator interacts with the drug-resistant variant(s) of the target protein, thereby rendering the drug-resistant variant susceptible to said drug.

15. The method of claim 13, wherein said modulator interacts with the drug-resistant variant(s) of the target protein, thereby rendering the drug-resistant variant susceptible to a new drug.

16. The method of claim 1, wherein the new drug is an antibiotic, an enzyme inhibitor, a protein inhibitor, an antibiotic resistance inhibitor, a herbicide resistance inhibitor, an insect repellant, an insecticide resistance inhibitor, a viral drug resistance inhibitor, a chemotherapeutic resistance inhibitor,

a modulator, an antagonist, an agonist, an effector, a ligand, an antibody, an antibody fragment, a peptide or a small molecule.

17. The method of claim 16, wherein the new drug is selected from a database of inhibitors or ligands.

18. The method of claim 16, wherein said database consists of small molecules.

19. The method of claim 16, wherein the new drug is a combination of at least two peptides or two small molecules.

20. The method of claim 1, wherein the new drug is designed using structure based drug design (SBDD), virtual ligand screening, ligand screening, high-throughput screening and ultra high-throughput screening.

21. The method of claim 1, wherein at least one drug-resistant variant has at least one mutation resulting in a structural change relative to the target protein.

22. The method of claim 1, wherein at least one drug-resistant variant has a combination of two or more mutations resulting in a structural change relative to the target protein.

23. The method of claim 1, wherein at least one drug-resistant variant has at least one mutation that results in a change of stability relative to the target protein.

24. The method of claim 1, wherein at least one drug-resistant variant has at least one global suppressor mutation that reverses the destabilizing effects of another antibiotic or inhibitor-resistant mutation.

25. The method of claim 1, wherein the target protein is a bacterial protein, viral protein, microbial protein, fungal protein, insect protein, cellular protein contributing to the development of cancer or cellular protein involved in inflammation or autoimmune disease.

26. The method of claim 1, wherein the target protein is from *Haemophilus influenzae*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, *Escherichia coli*, *Salmonella*, *Pseudomonas aeruginosa*, *Bacillus anthracis*, *Yersiniia pestis*, *Francisella tularensis*,

Brucella, Coxiella burnetii, staphylococci, Clostridium botulinum, Bordetella pertussis, Neisseria meningitidis Z2491, *Streptococcus mutan, Klebsiella pneumoniae, Chlamydia pneumoniae, Chlamydia trachomatis, Heliobacter pylori, Mycoplasma pneumoniae, Legionella pneumophila, Legionella micdadei, Treponema pallidum, Toxoplasma gondii, Borrelia burgdorferi, Campylobacter jejuni, Listeria monocytogenes, Neisseria gonnorhea, Neisseria meningitidis, Rickettsia prowazekii, Mycoplasma genitalium, Mycobacterium tuberculosis, Streptococcus pyogenes, Vibrio cholerae, Trypanosoma cruzi, Cryptosporidium parvum,* human immunodeficiency virus (HIV), cytomegalovirus (CMV), hepatitis C virus (HCV), herpes simplex virus (HSV), bovine diarrhea virus, human rhinovirus (HRV), hepatitis B virus (HBV), influenza virus, variola virus (smallpox), viral encephalitides, arenaviridae virus, bunyaviridae virus, filoviridae virus, flaviviridae virus, and filamentous fungi (Fusarium, Myrotecium, Cephalosporium, Trichoderma, Verticimonosporium, Stachybotrys species).

27. The method of claim 1, wherein the target protein is an enzyme, a transcription factor, a DNA polymerase, an RNA polymerase, a reverse transcriptase, a protease, a ribosomal protein or a DNA binding protein.

28. The method of claim 1, wherein the target protein is a Topo IV B subunit (ParE), a gyrase B subunit, a β-lactamase, a penicillinase, a cephalosporinase, a choramphenicol acetyltransferase, an aminoglycoside modifying enzyme, a D-ala D-ala ligase (ddl), a Peptide deformylase, an FabI, a GlmU, a MurA, a MurB, a MurC, a MurD, a FtsZ, a Dam methylase, an emr/mmr multidrug resistance protein, a bacterial DNA methylase, a squalene synthase, a human immunodeficiency virus reverse transcriptase, a human immunodeficiency virus protease, a human immunodeficiency virus integrase, a human immunodeficiency virus TAT/TAR protein, a human immunodeficiency virus NCp7 protein, a feline immunodeficiency virus protease, a CCR5/CXCR4 protein, a cytomegalovirus polymerase, an HCV NS3 helicase, an HCV NS3-4A serine protease, an HSV protease, an HSV TS protein, a Bovine diarrhea virus NS5B protein involved in RNA synthesis, an HRV 3C protease, an HBV polyphosphate kinase, an HMG Co-A reductase, a Chitin synthase-1, a D-glucan synthase, an inositol phosphorylceramide synthase, an EF-2 protein, a neuraminidase, a topoisomerase, a thrombin, a Factor Xa, a cathepsin B, C, S or K, a caspase, a matrix metalloproteinase, a cyclin-

dependent kinase, a urokinase, an EGFR tyrosine kinase, a Janus kinase family (JAK), a STAT protein, a Bcl-2 protein, a receptor family tyrosine kinase insulin receptor kinase, a protein-tyrosine phosphatase, a Src family kinase, a Lck family kinase, a ZAP/SYK family kinase, a TEC family tyrosine kinase, an angiotensin-converting enzyme, a monoamine oxidase, a tetrahydrofolate reductase or a p53 protein.

29. The method of claim 1, wherein the target protein is Beta-lactamase and the drug is a Beta-lactam antibiotic.

30. The method of claim 29, wherein the drug is cefotaxime.

31. The method of claim 1 that further comprises identifying a second new drug effective against at least one second generation drug-resistant variant comprising:

(a) selecting at least one second generation drug-resistant variant by exposing the drug-resistant variant of claim 1 to the new drug;

(b) obtaining the three dimensional structure of at least one second generation site of interaction with the new drug;

(c) comparing the three dimensional structures of the second generation variant and the variant sites of interaction with the drug to identify structural similarities and/or differences; and

(d) designing a second new drug to interact with the second generation variant by using the structural similarities or differences.

32. The method of claim 1, wherein SAR or another structure/function-based algorithm is used to design the new drug.

33. The method of claim 1, wherein said at least one drug-resistant variant(s) is two to nine variants.

34. The method of claim 1, wherein said at least one drug-resistant variant(s) is about ten or more variants.

35. The method of claim 1, wherein said at least one drug-resistant variant(s) is about 100 or more variants.

36. The method of claim 1, wherein said at least one drug-resistant variant(s) is about 1000 or more variants.

37. A method for preparing a library representing a drug resistance evolution profile of a target protein or target protein fragment comprising:

(a) generating a first population of sequence or structure variants of said target protein by one or more rounds of a first method of directed evolution in the presence of said drug;

(b) generating at least one other population of sequence or structure variants by one or more further rounds of said first method of directed evolution or by at least one other method of directed evolution or by obtaining a library of natural variant isolates;

(c) comparing sequences or structures or fragments of said variants in said first population to the sequences or structures or fragments of said variants in said at least one other population to identify variants having unique sequences or structures; and

(d) preparing a library comprising said unique sequence variants that represents a drug resistance evolution profile of said target protein or target protein fragment.

38. A library composed of drug resistant variants obtained by the method of claim 37, in which the library is profiled by primary amino acid sequence.

39. A library composed of drug resistant variants obtained by the method of claim 37, in which the library is profiled by three-dimensional structure.

40. A data structure for storing information on directed evolution of a series of drug resistant variants comprising:

(a) a series of separate variant data fields generated by directed evolution, each variant data field comprising each x, y and z atomic coordinate of each drug resistant variant;

(b) a target protein data field comprising each x, y and z atomic coordinate of the target protein;

(c) a structural difference data field comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein; and

(d) a series of drug resistance data fields, each drug resistance data field comprising each x, y and z atomic coordinate of each drug that is resistant to a drug resistance variant.

41. A method for identifying a new drug effective against at least one drug-resistant variant of a target protein comprising screening the library produced by the method of claim 37 with at least one new drug.

42. A method for identifying a new drug effective against more than one drug-resistant variant of a target protein comprising screening the library produced by the method of claim 37 with at least one new drug.

43. A new drug identified by the method of claim 40 or 41.

44. A method for identifying regions in a target protein that are structurally altered within a drug-resistant variant of the target protein, comprising comparing the structures of the variants of the drug resistance evolution profile produced by the method of claim 37.

45. A library composed of drug resistant variants produced by the method of claim 37, wherein said library is composed of at least 10 variants.

46. The library of claim 44, wherein said library is composed of at least 100 variants.

47. The library of claim 44, wherein said library is composed of at least 1000 variants.

48. A computer-assisted method for identifying a new drug effective against at least one drug-resistant variant of a target protein, comprising:

(a) obtaining the three dimensional structure of at least one variant site of interaction with the drug by using computer processing to generate an x, y and z coordinate for each atom in the variant site of interaction;

(b) obtaining the three dimensional structure of the target protein site of interaction with the drug by using computer processing to

generate an x, y and z coordinate for each atom in the target protein site of interaction;

(c) measuring the differences in distance between the x, y and z coordinates for each atom in the variant and target protein sites of interaction;

(d) designing a new drug to interact with the site of interaction of at least one of the variants by using computer processing to fit a library of potential new drugs into the variant and target protein sites of interaction and determining which potential new drug fits into the variant site of interaction better than into the target protein site of interaction.

49. An article of manufacture having instructions stored on it that cause a computing system to identify a new drug effective against at least one drug-resistant variant of a target protein, the instructions comprising:

(a) instructions for comparing a series of separate variant data fields with a target protein data field to generate a structural difference data field comprising the differences in x, y and z atomic coordinates for each drug resistant variant relative to the x, y and z atomic coordinates of the target protein, wherein each variant data field within the series of separate variant data fields comprises each x, y and z atomic coordinate of each drug resistant variant, and wherein a target protein data field comprises each x, y and z atomic coordinate of the target protein;

(b) instructions for generating a new drug data field from the structural difference data field and a small molecule data field, wherein the small molecule data field comprises x, y and z atomic coordinates for a series of small molecules and wherein the new drug data field comprises small molecules that can optimally fit within the site of interaction of the drug-resistant variant.

50. A new drug identified using the article of manufacture of claim 49.

51. A method for identifying a new antibiotic effective against at least one antibiotic-resistant of bacteria, comprising:

(a) selecting at least one antibiotic-resistant strain of bacteria by exposing the bacteria to a selecting antibiotic;

(b) obtaining the three dimensional structure of a target protein site of interaction with the antibiotic, wherein the target protein is within the bacteria and wherein the antibiotic has its effect through the target protein;

(c) obtaining the three dimensional structure of a target protein site of interaction with the antibiotic, wherein the target protein is within the bacteria and wherein the antibiotic has its effect through the target protein;

(d) obtaining the three dimensional structure of a variant protein site of interaction with the antibiotic, wherein the variant protein is within at least one antibiotic-resistant strain of bacteria selected in step (a) and wherein the variant protein is resistant to the selecting antibiotic;

(e) comparing the three dimensional structures of the target and variant sites of interaction to identify structural similarities and differences; and

(f) designing a new drug to interact with at least one of the variants by using the structural similarities or differences.

52. A system, comprising:

a computer system;

memory accessible to the computer system;

a first data structure stored in the memory comprising three dimensional structural data of a target protein;

a second data structure stored in the memory comprising a first population of three dimensional structural data for a series of variants of the target protein, wherein the variants were generated by one or more rounds of directed evolution of the target protein in the presence of a drug;

a third data structure stored in the memory comprising a second population of three dimensional structural data for a series of second generation variants of a first variant, wherein the second generation of variants is generated by one or more rounds of directed evolution of the first variant in the presence of a drug;

a fourth data structure comprising three dimensional structural data for a population of potential new drugs comprising antibiotics, enzyme inhibitors, protein inhibitors, antibiotic resistance inhibitors, herbicide resistance inhibitors, insect repellants, insecticide resistance inhibitors, viral drug resistance inhibitors, chemotherapeutic resistance inhibitors, modulators, antagonists, agonists, effectors, ligands, antibodies, antibody fragments, peptides or small molecules;

a first software component operable on the computer system to compare the first data structure with the second or third data structure and to identify three dimensional structural variants that represent a drug resistance evolution profile of said target protein;

a second software component operable on the computer system to compare the fourth data structure with the drug resistance evolution profile of said target protein and thereby to design a new drug effective against at least one drug resistance variant.

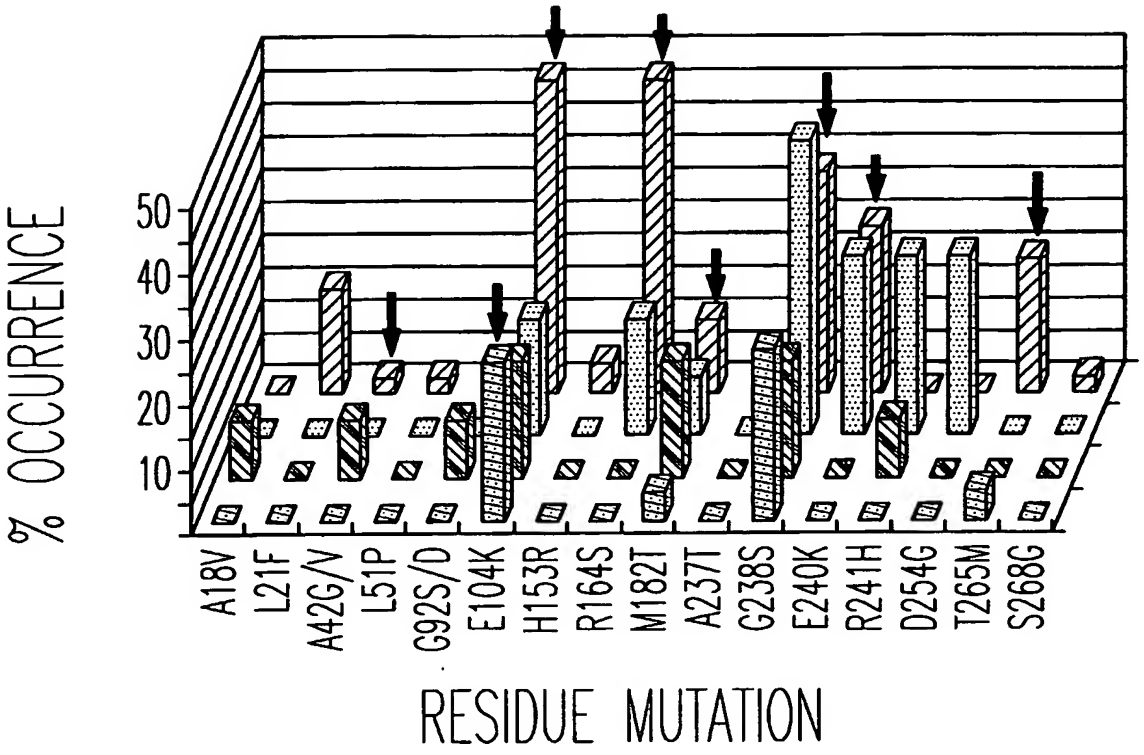

% OCCURRENCE
50
40
30
20
10
A18V
L21F
A42G/V
L51P
G92S/D
E104K
H153R
R164S
M182T
A237T
G238S
E240K
R241H
D254G
T265M
S268G
RESIDUE MUTATION
EXPERIMENT
NATURAL MUTANTS
DNA ANALOGS
GENE SHUFFLING
MUTATOR STRAIN

Fig. 1

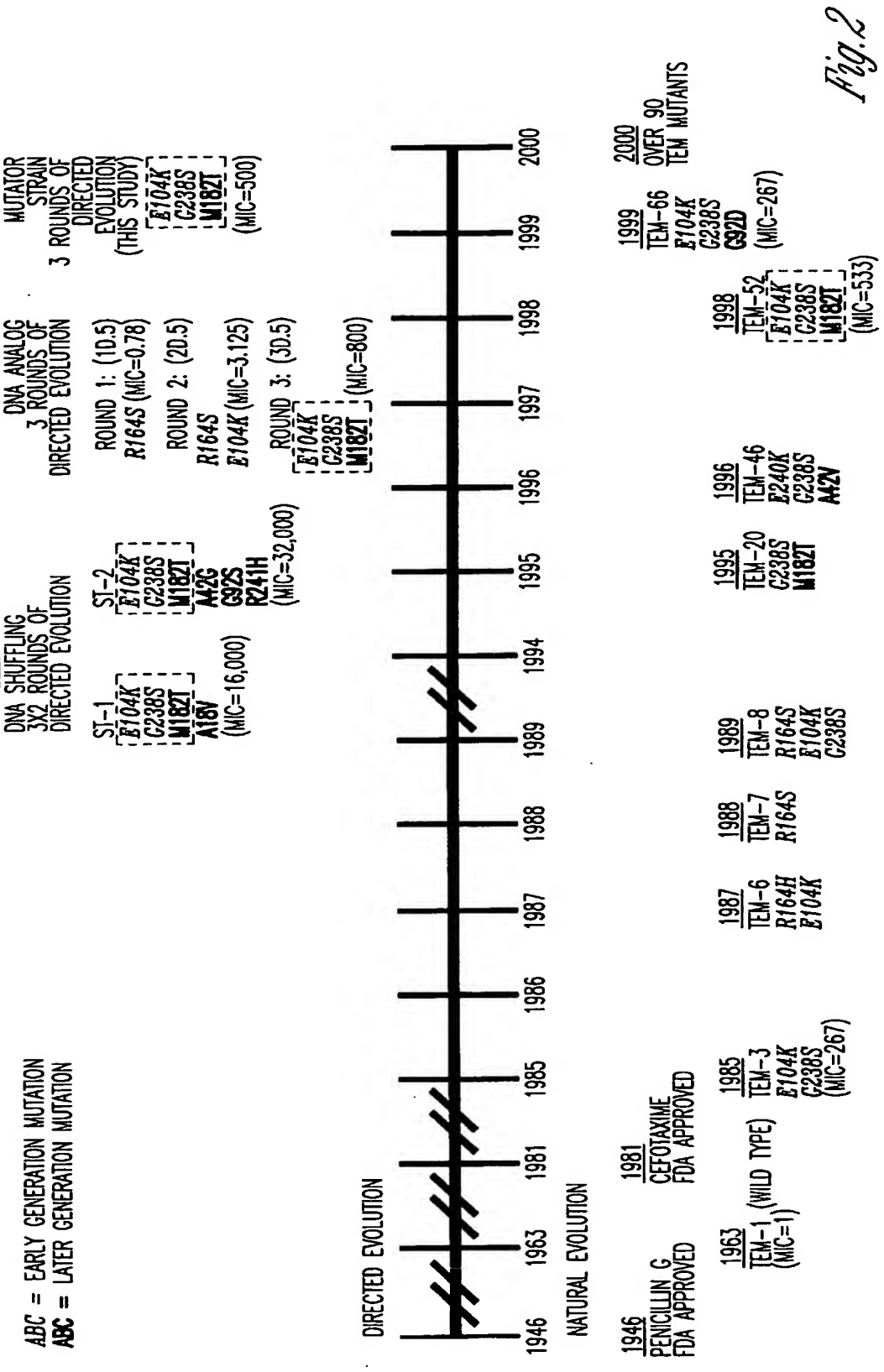

ABC = EARLY GENERATION MUTATION
ABC = LATER GENERATION MUTATION
DIRECTED EVOLUTION
NATURAL EVOLUTION
1946
PENICILLIN G
FDA APPROVED
1963
TEM-1 (WILD TYPE)
(MIC=1)
1981
CEFOTAXIME
FDA APPROVED
1985
TEM-3
E104K
G238S
(MIC=267)
1987
TEM-6
R164H
E104K
1988
TEM-7
R164S
1989
TEM-8
R164S
E104K
G238S
1994
DNA SHUFFLING
3X2 ROUNDS OF
DIRECTED EVOLUTION
ST-1
E104K
G238S
M182T
A18V
(MIC=16,000)
ST-2
E104K
G238S
M182T
A42G
G92S
R241H
(MIC=32,000)
1995
TEM-20
G238S
M182T
1996
TEM-46
E240K
G238S
A42V
1998
DNA ANALOG
3 ROUNDS OF
DIRECTED EVOLUTION
ROUND 1: (10.5)
R164S (MIC=0.78)
ROUND 2: (20.5)
R164S
E104K (MIC=3.125)
ROUND 3: (30.5)
E104K
G238S
M182T (MIC=800)
1998
TEM-52
E104K
G238S
M182T
(MIC=533)
1999
TEM-66
E104K
G238S
G92D
(MIC=267)
2000
MUTATOR
STRAIN
3 ROUNDS OF
DIRECTED
EVOLUTION
(THIS STUDY)
E104K
G238S
M182T
(MIC=500)
2000
OVER 90
TEM MUTANTS
1946
1963
1981
1985
1986
1987
1988
1989
1994
1995
1996
1997
1998
1999
2000

Fig.2

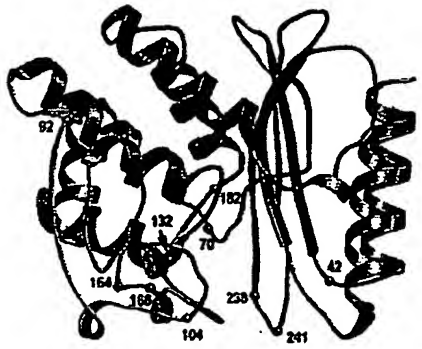


Fig. 3A

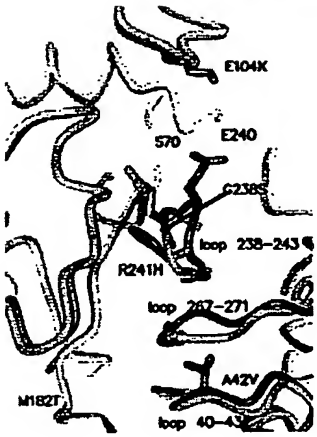


Fig. 3B

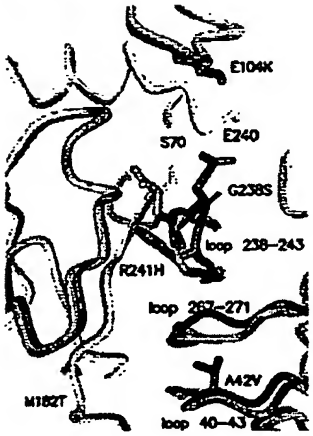


Fig. 3C

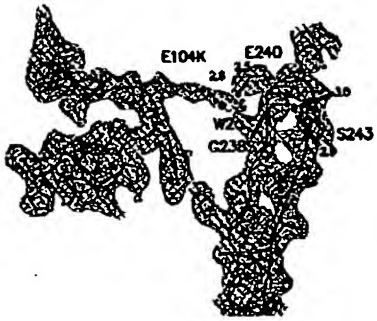


Fig. 3D

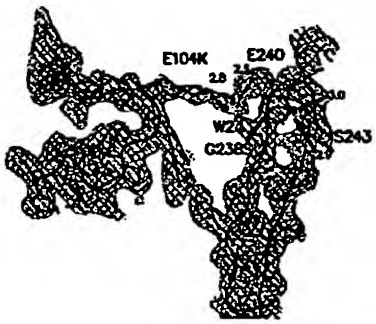


Fig. 3E

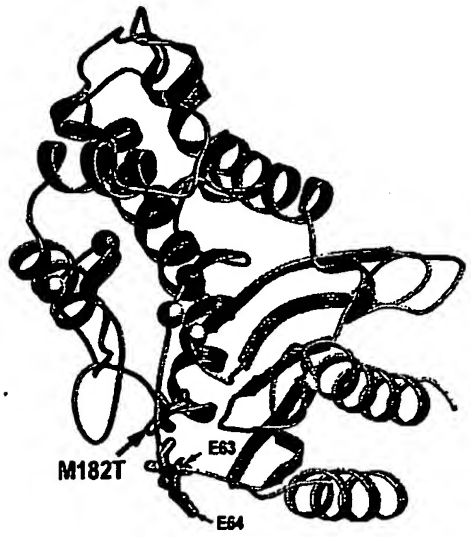


Fig. 4A

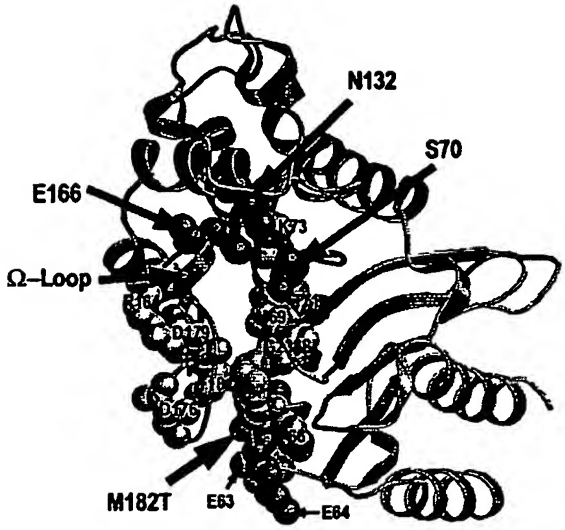


Fig. 4B

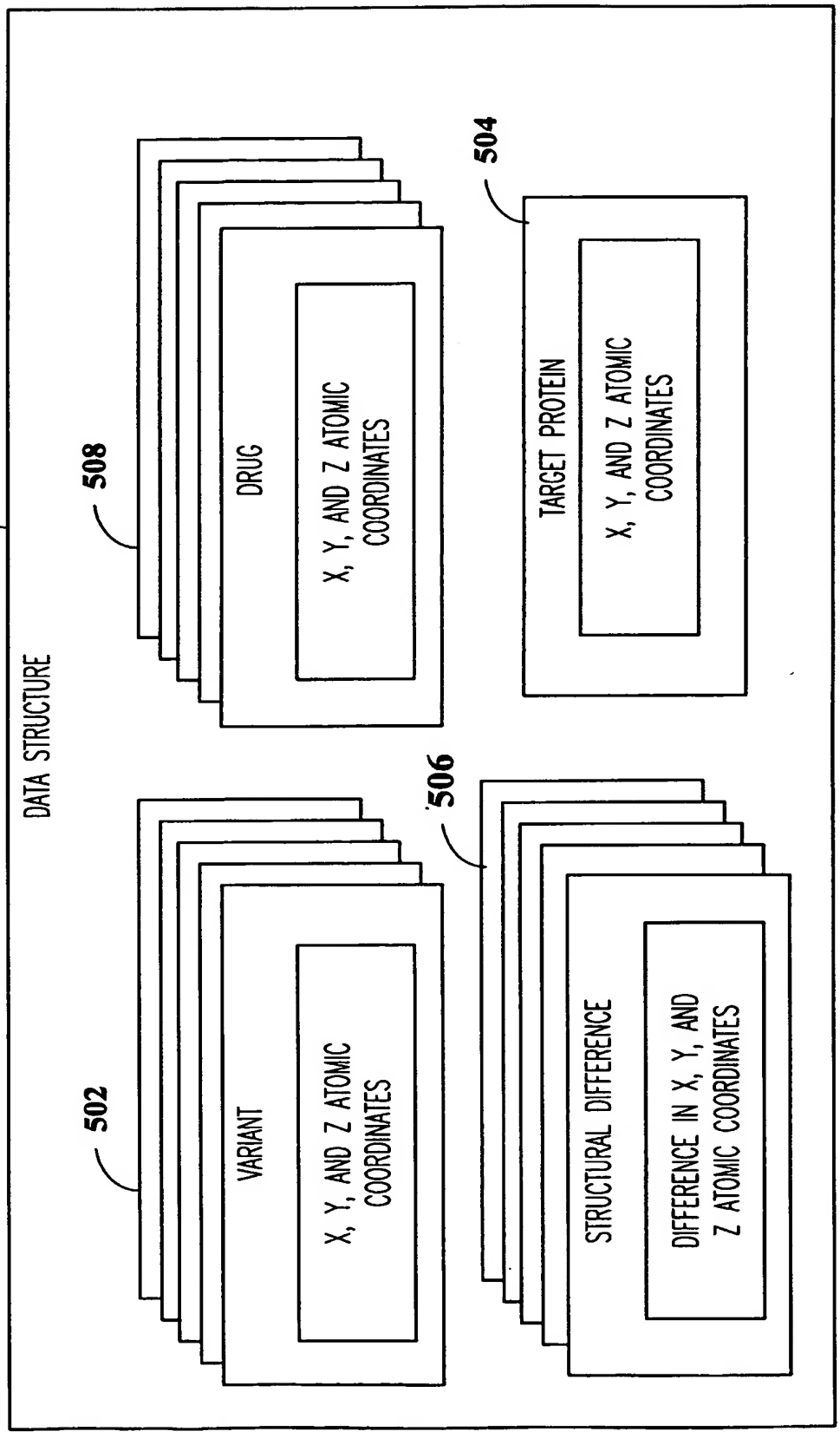

500
DATA STRUCTURE
502
VARIANT
X, Y, AND Z ATOMIC COORDINATES
504
TARGET PROTEIN
X, Y, AND Z ATOMIC COORDINATES
506
STRUCTURAL DIFFERENCE
DIFFERENCE IN X, Y, AND Z ATOMIC COORDINATES
508
DRUG
X, Y, AND Z ATOMIC COORDINATES

Fig. 5

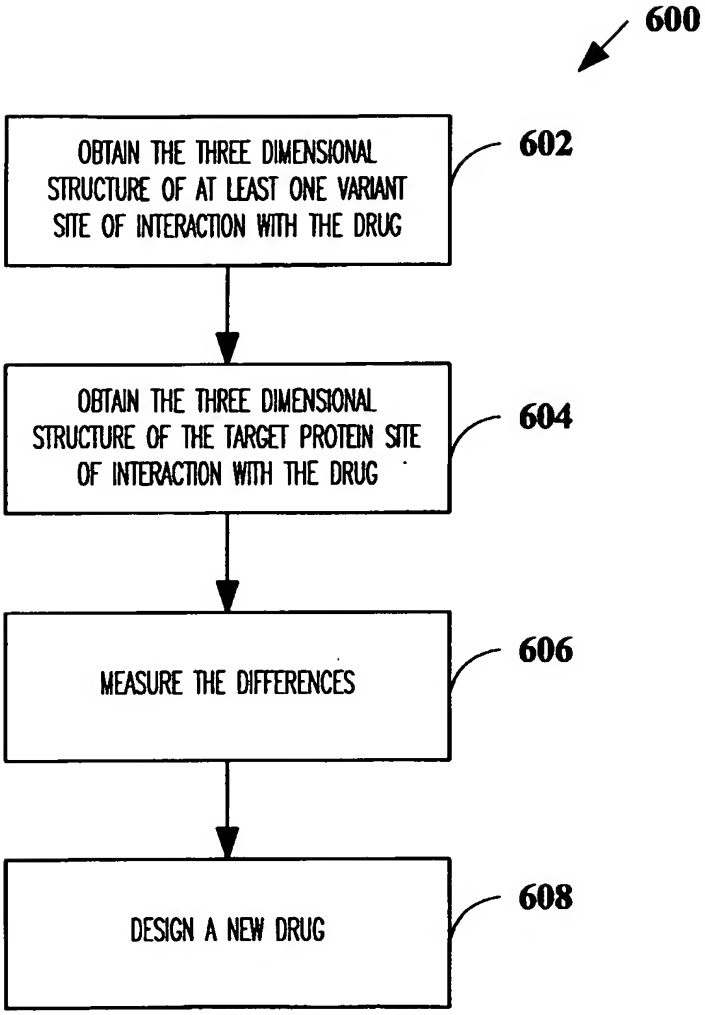

600
OBTAIN THE THREE DIMENSIONAL STRUCTURE OF AT LEAST ONE VARIANT SITE OF INTERACTION WITH THE DRUG
602
OBTAIN THE THREE DIMENSIONAL STRUCTURE OF THE TARGET PROTEIN SITE OF INTERACTION WITH THE DRUG
604
MEASURE THE DIFFERENCES
606
DESIGN A NEW DRUG
608

Fig. 6

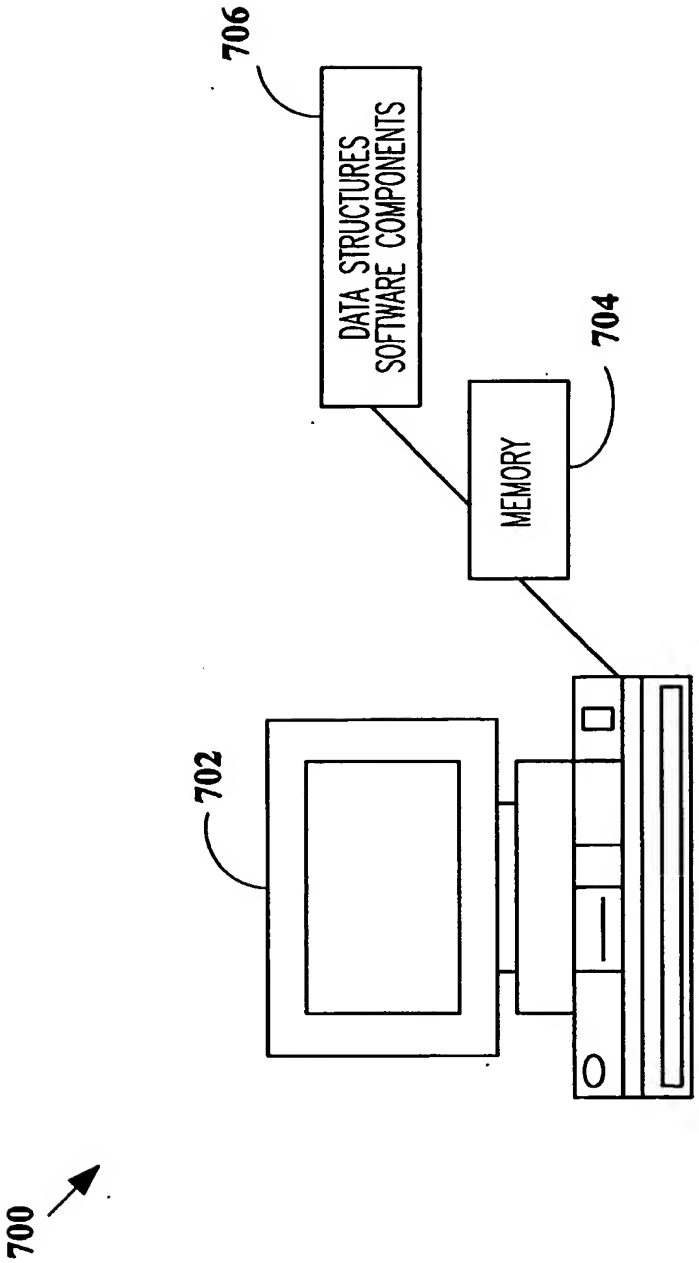

706
DATA STRUCTURES
SOFTWARE COMPONENTS
704
MEMORY
702
700

Fig.2

SEQUENCE LISTING

<110> The Scripps Research Institute
Stevens, Raymond C.
Orencia, Maria Celilia

<120> Small molecule design and use of crystallographic structures for development of new antibiotics

<130> 1361.006WO1

<150> US 60/272,248
<151> 2001-02-28

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> A primer

<400> 1
atgagtattc aacatatgtt cctggtcgcc 30

<210> 2
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> A primer

<400> 2
ttaccaatgc ttggatccaa tcagtgaggc 30

<210> 3
<211> 861
<212> DNA
<213> Klebsiella pneumoniae

<400> 3
atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct 60
gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca 120
cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc 180
gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc 240
cgtgttgacg ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg 300
gttaagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta 360
tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct gacaacgatc 420
ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt 480
gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgacg 540
cctgcagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct 600
tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc 660
tcggcccttc cggctggctg gtttattgct gataaatctg gagccagtga gcgtgggtct 720
cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac 780
acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga gataggtgcc 840
tcactgatta agcattggta a 861

<210> 4
<211> 286
<212> PRT
<213> Klebsiella pneumoniae

<400> 4
Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
1 5 10 15
Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys
20 25 30
Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
35 40 45
Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
50 55 60
Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser
65 70 75 80
Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
85 90 95

Gln Asn Asp Leu Val Lys Tyr Ser Pro Val Thr Glu Lys His Leu Thr
100 105 110
Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser
115 120 125
Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys
130 135 140
Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
145 150 155 160
Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
165 170 175
Asp Thr Thr Thr Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
180 185 190
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
195 200 205
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
210 215 220
Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Ser Glu Arg Gly Ser
225 230 235 240
Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
245 250 255
Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
260 265 270
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
275 280 285